PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-074210
(43)Date of publication of application : 17.03.1998

(51)Int.Cl. G06F 17/30

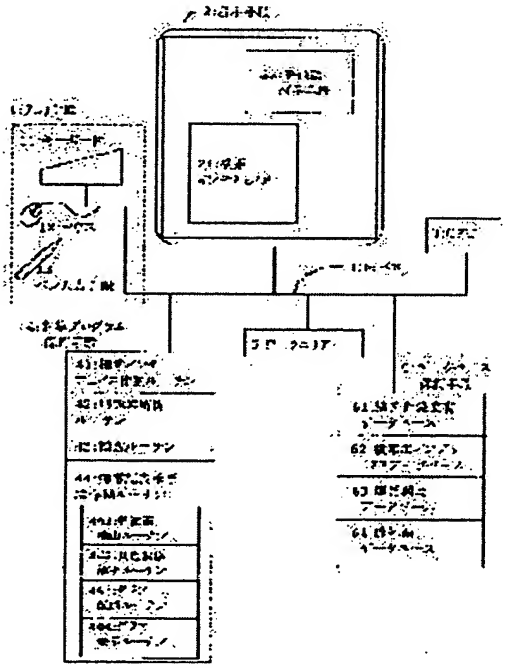
(21)Application number : 09-178500
(22)Date of filing : 03.07.1997

(71)Applicant : HITACHI LTD
(72)Inventor : NIWA YOSHIKI
SAKURAI HIROBUMI

(30)Priority
Priority number : 08176174 Priority date : 05.07.1996 Priority country : JP

(54) METHOD AND DEVICE FOR SUPPORTING DOCUMENT RETRIEVAL AND DOCUMENT RETRIEVING SERVICE USING THE METHOD AND DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a retrieving method for enabling a user to have a look at the whole image of a retrieved document group and to attain retrieval as service.
SOLUTION: A feature word displaying means 22 is displayed on a display means 2, a word group characteristically appearing in a document group retrieved by a user's request is extracted, mutual relation among feature words is checked, a graph setting the feature words as nodes is prepared, and the whole image of retrieved results is displayed on the means 22. When a user selects his (or her) interested word or uninterested word while observing the displayed feature word graph, succeeding retrieval strategy can be effectively prepared.



LEGAL STATUS

[Date of request for examination] 22.11.2001
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number] 3614618
[Date of registration] 12.11.2004

(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（Ａ）　(11)特許出願公開番号

特開平10－74210

(43)公開日　平成10年(1998)３月17日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ　17/30			Ｇ０６Ｆ　15/403	３４０Ｂ	
			15/401	３２０Ｚ	

審査請求　未請求　請求項の数17　ＯＬ　（全 27 頁）

(21)出願番号　特願平9－178500

(22)出願日　平成９年(1997)７月３日

(31)優先権主張番号　特願平8－176174

(32)優先日　平８(1996)７月５日

(33)優先権主張国　日本（ＪＰ）

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　丹羽　芳樹
埼玉県比企郡鳩山町赤沼2520番地　株式会社日立製作所基礎研究所内

(72)発明者　櫻井　博文
埼玉県比企郡鳩山町赤沼2520番地　株式会社日立製作所基礎研究所内

(74)代理人　弁理士　高橋　明夫　（外１名）

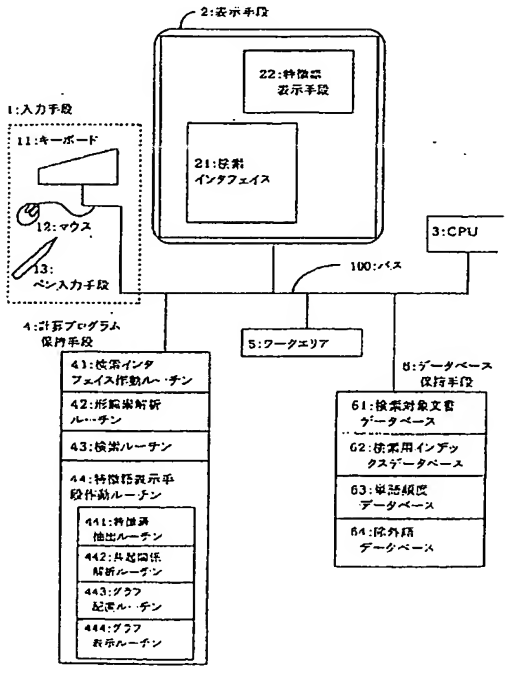
(54)【発明の名称】　文献検索支援方法及び装置およびこれを用いた文献検索サービス

(57)【要約】

【課題】　検索された文書群の全体像を一覧することが可能な検索方法を提供すること。またサービスとしての検索を可能とすること。

【解決手段】　表示手段に特徴語表示手段を表示し、ユーザーからの要求により検索された文書群に特徴的に出現する語群を抽出し、さら特徴語相互間の関連性を調べて、特徴語をノードとするグラフを作成し検索結果の全体像を特徴語表示手段に表示する。さらに、ユーザーは表示された特徴語のグラフを見て、自分の関心の強い語や逆に関心のない語を選択することにより、効果的に次の検索戦略を立てられるようになる。

図１



1

【特許請求の範囲】

【請求項1】設定されたキーワードに応じて検索対象文書群から前記キーワードを持つ文書を検索結果文書として検出すること、ある単語が前記検索結果文書群中のいくつの文書に現れるかを意味する単語の文書頻度を検出すること、前記単語が検索対象文書群全体においていくつの文書に出現するかを意味する単語の全体文書頻度を検出すること、前記単語の文書頻度と単語の全体文書頻度との比を意味する頻度比を導出すること、前記文書頻度を所定の関係で頻度クラスに区分分けして各単語の文書頻度に応じて各単語を頻度クラスに対応させること、各頻度クラスから適当数の単語を単語の頻度比の大きさ順に特徴語として抽出すること、抽出された特徴語をグラフ形式またはリスト形式で表示することとよりなることを特徴とする文献検索支援方法。

【請求項2】前記抽出された特徴語を頻度クラス別のリスト形式または特徴語間の関連を示すグラフ形式のいずれかで表示する請求項1記載の文献検索支援方法。

【請求項3】設定されたキーワードに応じて検索対象文書群から前記キーワードを持つ文書を検索結果文書として検出する手段、ある単語が前記検索結果文書群中のいくつの文書に現れるかを意味する単語の文書頻度を検出する手段、前記単語が検索対象文書群全体においていくつの文書に出現するかを意味する単語の全体文書頻度を検出する手段、前記単語の文書頻度と単語の全体文書頻度との比を意味する頻度比を導出する手段、前記頻度比を所定の関係で頻度クラスに区分分けして各単語の頻度比に応じて各単語を頻度クラスに対応させる手段、各頻度クラスから適当数の単語を単語の頻度比の大きさ順に特徴語として抽出する手段、抽出された特徴語をグラフ形式またはリスト形式で表示する手段とよりなることを特徴とする文献検索装置。

【請求項4】前記抽出された特徴語を頻度クラス別のリスト形式または特徴語間の関連を示すグラフ形式のいずれかで表示するする手段および特徴語表示形式を選択指定する手段を有する請求項3記載の文献検索装置。

【請求項5】前記特徴語間の関連が特徴語間の共起関係を基礎として決定され、前記グラフ形式が特徴語をノードとし関連性の高い特徴語の単語対にリンクを張って構成されたグラフである請求項3または4記載の文献検索装置。

【請求項6】設定されるキーワードが必須キーワード、加点キーワードおよび減点キーワードの3種類のキーワードとされ、必須キーワードによる検索は各必須キーワードによるアンド条件で検索を行ない、検索された前記検索結果文書群の各文書について、加点キーワードを含む場合には加点キーワード数に応じて高い得点を与え、減点キーワードを含む場合には減点キーワード数に応じて減点し、より高い得点を得た文書群から特徴語抽出を行なう請求項3または4記載の文献検索装置。

2

【請求項7】必須キーワードの設定のない場合に加点キーワードによる検索が行われ、各加点キーワードによる検索はオア条件で行われる請求項6記載の文献検索装置。

【請求項8】設定される必須キーワード、加点キーワードおよび減点キーワードの3種類のキーワード間でキーワードの種類を変更可能とされるとともに、表示された特徴語を必須キーワード、加点キーワードおよび減点キーワードのいずれかに複写可能とした請求項6または7記載の文献検索装置。

【請求項9】特徴語のグラフ表示において縦軸方向が検索された文書群における特徴語の文書頻度を表す請求項4ないし8のいずれかに記載の文献検索装置。

【請求項10】検索元から伝送されたキーワードに応じて検索対象文書群から前記キーワードを持つ文書を検索結果文書として検出すること、ある単語が前記検索結果文書群中のいくつの文書に現れるかを意味する単語の文書頻度を検出すること、前記単語が検索対象文書群全体においていくつの文書に出現するかを意味する単語の全体文書頻度を検出すること、前記単語の文書頻度と単語の全体文書頻度との比を意味する頻度比を導出すること、前記頻度比を所定の関係で頻度クラスに区分分けして各単語の頻度比に応じて各単語を頻度クラスに対応させること、各頻度クラスから適当数の単語を単語の頻度比の大きさ順に特徴語として抽出すること、抽出された特徴語を特徴語間の関連を示すグラフ形式で表示可能なデータとして構成することまたは抽出された特徴語を頻度クラス別のリスト形式で表示可能なデータとして構成すること、前記特徴語をグラフ形式またはリスト形式で表示可能なデータとして検索元に送信することよりなる文献検索サービス方法。

【請求項11】前記検索元は、少なくとも、抽出すべきキーワードを持つ文書を特定するためのキーワードを伝送するための手段および前記送信された特徴語および特徴語間の関連を示すグラフ形式またはリスト形式で表示可能なデータを受信して表示するための手段を備えて検索サービスを受ける請求項10記載の文献検索サービス方法。

【請求項12】前記検索元は、前記送信された特徴語および特徴語間の関連を示すグラフ形式またはリスト形式で表示可能なデータを表示ソフトとともに伝送されて検索サービスを受ける請求項10記載の文献検索サービス方法。

【請求項13】前記検索元は、検索さービスを受けるためのユーザインタフェイス駆動ソフトを検索作業の開始時あるいは前もって検索サービス提供者から伝送を受けこれを駆動して検索サービスを受ける請求項10記載の文献検索サービス方法。

【請求項14】検索結果に出現する各語の特徴度を計算するための頻度データを記録したコンピュータ読み取り

3

可能な記録媒体であって、各語に関するデータが、
（ａ）文字列、（ｂ）検索された文書の内の何件にその語が出現したかを表す文書頻度、（ｃ）検索結果に関係なく、検索対象文書全体で何件の文書に使われているかを表すデータベース全体での文書頻度、（ｄ）前記検索結果における文書頻度とデータベース全体での全体文書頻度から計算される検索結果におけるその語の特徴度、（ｅ）前記検索結果における文書頻度の大小によってクラス分けした場合の頻度クラスとからなり、前記頻度クラスのそれぞれから前記特徴度の上位にある語を検索対象文書群における特徴語とすることを特徴とする検索結果に出現する語の頻度データを記録したコンピュータ読み取り可能な記録媒体。

【請求項１５】検索結果に出現する特徴語間の関連度を計算するために、特徴語が共出現する共起データを記録したコンピュータ読み取り可能な記録媒体であって、各特徴語対に関するデータが、（ａ）検索結果文書群における両特徴語が共出現する共起頻度と（ｂ）該共起頻度と両特徴語各々の検索結果に出現する頻度データから計算される両特徴語の関連度とからなり、前記関連度の高い特徴語対に関連性が強いことを示すリンクを張れるようにすることを特徴とする検索結果における特徴語間の共起データを記録したコンピュータ読み取り可能な記録媒体。

【請求項１６】検索結果に出現する特徴語対のグラフを画面表示するためのデータを記録したコンピュータ読み取り可能な記録媒体であって、前記特徴語対のグラフを画面表示するためのデータは（ａ）グラフのノード部分に特徴語を表示するためのデータ、（ｂ）特徴語間の関連性を示すリンクを表示するためのデータとからなるとともに、前記各ノードのデータは、中心座標、表示する文字列、および、文字列を表示する領域の縦横の文字数とサイズからなり、前記各リンクのデータは始点座標と終点座標とからなり、特徴語グラフをリンクと文字列とによる二次元表示を可能としたことを特徴とする特徴語グラフを画面表示するためのデータを記録したコンピュータ読み取り可能な記録媒体。

【請求項１７】検索結果に出現する各語の特徴度を計算し、特徴語を導出し、特徴語対の共出現頻度にもとづいて関連性の高いと判定される特徴語対にリンクを張って得られる特徴語のグラフを画面表示するデータを記録したコンピュータ読み取り可能な記録媒体であって、前記検索結果に出現する各語に関するデータが、（ａ）文字列、（ｂ）検索された文書の内の何件にその語が出現したかを表す文書頻度、（ｃ）検索結果に関係なく、検索対象文書全体で何件の文書に使われているかを表すデータベース全体での文書頻度、（ｄ）前記検索結果における文書頻度とデータベース全体での全体文書頻度から計算される検索結果におけるその語の特徴度、（ｅ）前記検索結果における文書頻度の大小によってクラス分けした場合の頻度クラスとからなり、前記頻度クラスのそれぞれから前記特徴度の上位にある語を検索対象文書群における特徴語とし、
前記特徴語間の関連度を計算するために、各特徴語対に関するデータが、（ｆ）検索結果文書群における両特徴語が共出現する共起頻度と（ｇ）該共起頻度と両特徴語各々の検索結果に出現する頻度データから計算される両特徴語の関連度とからなり、前記関連度の高い特徴語対にリンクを張れるようにし、
前記リンクの張られた特徴語グラフを画面表示するために、前記特徴語グラフを画面表示するためのデータは（ｈ）グラフのノード部分に特徴語を表示するためのデータ、（ｉ）特徴語間の関連性を示すリンクを表示するためのデータとからなるとともに、前記各ノードのデータは、中心座標、表示する文字列、および、文字列を表示する領域の縦横の文字数とサイズからなり、前記各リンクのデータは始点座標と終点座標とからなり、前記各特徴語グラフをリンクと文字列とによる二次元表示を可能としたことを特徴とする特徴語グラフを画面表示するためのデータを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、文献検索における対話的なガイダンス機能を実現するためのユーザインタフェイスを持つ文献検索支援方法及び装置およびこれを用いた文献検索サービスに関する。

【０００２】

【従来の技術】文献検索においては、ユーザーが所望する文献集合に早く容易に到達できるように、文献検索装置とユーザーとのさまざまなインタフェイスが考案、開発されている。その中の主なものとしてはフィードバックとガイダンスがある。フィードバックとは検索結果のいくつかのアイテムに対してユーザーが「当たり／はずれ」の判定を下すと、その判定を反映した検索結果を得ることができるしくみである。またガイダンスとは検索作業の各段階でその検索作業と関連のあると思われる情報、したがって利用者が検索条件を工夫したり改良したりするのに参考となると思われる情報を提供する機能である。

【０００３】ガイダンス機能については、従来一般に、入力された検索条件に対してその関連情報を提示する方法が行われている。例えば、シソーラスなど単語間の関連性を示すデータベースを保持しておき、検索条件として入力された語と関連のある語をデータベースから取り出して提示する方法である。シソーラスの場合には主に単語間の上位－下位関係を示す木構造のデータであるが、共起統計を用いて関連語データを自動生成しそれを用いる方法もある（例えば、B. R. Schatz et al, Interactive term suggestion for users of digital libra

5

ries: Using subject thesauri and co-occurrence lists for information retrieval. Proc. ACM DL96.p.126-133）。また、単語間の共起統計データに基づき検索語とその関連語をネットワーク状に表示する方法も提案されている（例えば、R.H. Fowler, D. W. Dearholt, Information Retrieval Using Pathfinder Networks. In Pathfinder Associative Networks, Ablex, article 12, Edited by R. W. Schvaneveldt(1990)）。

【０００４】しかしながら、検索条件に対してその関連情報を提示する方法では、検索語が複数になった場合や否定が使われた場合の対処が難しく、またキーワードを用いない書類の検索（連想検索など）にも対処が難しいという問題がある。これを克服する方法として、検索結果から関連情報を自動抽出してユーザに提供する手法がある。例えば、スキャター・ギャザー法（D.Cutting他(1992). Scatter/Gather : A Cluster-based Approach to Browsing Large Document Collections. Proc. ACM SIGIR'92,p.318-329）では検索された文書群を自動分類（クラスタリング）して各クラスごとの特徴語を表示するものである。しかし、クラスタリングは文書数が増えると計算量が2乗あるいは3乗のオーダーで大きくなるのでリアルタイムでの反応が難しくなり、また一般に検索作業が進んで行くとクラス間の違いが微妙になり、クラスの特徴語からそのクラスの性格を把握しにくくなるという問題があった。

【０００５】

【発明が解決しようとする課題】本発明は、前述の問題を解消して、検索された文書群に含まれる話題群をリアルタイムで一覧できるよう文書群に特徴的に現れる語群の特徴語をグラフ形式またはリスト形式で画面表示すること、さらには、文書群に特徴的に現れる語群を低頻度語から高頻度語までバランス良く抽出することのできる文献検索支援方法及び装置、さらには、この文献検索を希望するユーザが遠隔地からも行えるようにすることを目的とする。

【０００６】

【課題を解決するための手段】このため、検索された文書群に含まれる話題群をリアルタイムで一覧できるよう、文書群に特徴的に出現する語群をノードとし、さらに特徴語間に強い共起関係がある場合、すなわち同一文書中に出現しやすい度合が高い場合、その単語対にリンクを張ることによりグラフを構成し、そのグラフを画面表示するとともに、特徴語のグラフ表示の際に、一般的な語と特殊性の高い語を一目で見分けることができるように縦軸方向が特徴語の文書頻度を表すようにする。リストの例で言えば、特徴語を頻度クラスで分類し、文書頻度の高いものを上段に配列して一覧できるようにして特殊性の高い語を一目で見分けることができるようにする。検索された文書群から特徴語を選ぶ際に、低頻度の語から高頻度の語までバランス良く特徴語を抽出するためには、特徴語を出現頻度によってクラス分けを行ない、それぞれのクラスから頻度比、すなわち当該文書群における文書頻度と検索対象全体における文書頻度の比が大きいものから順に抽出する。

6

【０００７】

【発明の実施の形態】

実施例1

以下、本発明の第1の実施例を図1－２０に従って説明する。本実施例は、独立に使用されるコンピュータによる検索装置の構成例である。本実施例では、検索結果をグラフ表示とする場合を主体に説明する。図1に本実施例の文献検索装置の全体構成を示す。１は入力手段、２は表示手段、３はＣＰＵ、４は計算プログラム保持手段、５は計算プログラムを動作させるためのワークエリア、６はデータベース保持手段であり、これらの手段あるいは装置は、これらの間で相互に信号のやり取りをするためのバス１００で連携される。入力手段1はキーボード１１、マウス１２、ペン入力手段１３などから構成され、表示手段2には検索インタフェイス２１および検索をガイドするための特徴語表示手段２２が表示される。計算プログラム保持手段4には本実施例の文献検索装置を動作させるために必要となる検索インタフェイス作動ルーチン４１、形態素解析ルーチン４２、検索ルーチン４３および特徴語表示手段作動ルーチン４４が格納される。特徴語表示手段作動ルーチン４４は検索された文書群から特徴語を抽出して特徴語表示手段２２に表示するため、特徴語抽出ルーチン４４１、共起関係解析ルーチン４４２、グラフ配置ルーチン４４３、およびグラフ表示ルーチン４４４をサブルーチンとして用いる。ワークエリア5についての詳細は図2を参照して後述する。データベース保持手段6は検索対象となる文書データベース６１、検索に用いるインデックスデータベース６２、単語頻度に関するデータベース６３および除外語データベース６４から構成される。これらのデータベースは、一般には、事前に準備されているものの中から、使用者が自分の検索目的に合うものを検索対象データとして選択して使用する。たとえば、新聞記事についての検索をしたいときは、新聞社が発行しているデータベースを購入して使用することになる。もっとも、除外語データベース６４は装置の供給者が付属データとして提供するものである場合があろう。

【０００８】図2はワークエリア５の構成についての詳細である。ワークエリア5は計算プログラム保持手段4にある諸ルーチンが動作するために必要となるパラメータや一時的なデータを保持するためのエリアであり、検索ワークエリア５１、特徴語抽出ワークエリア５２；共起関係解析ワークエリア５３、グラフ配置ワークエリア５４から構成される。各エリアには、更に細分されたデータエリアが備えられるが、これらの詳細についてはそれぞれの関連するルーチンが動作する時に説明する。ユ

7

ーザが文献検索をしようとするとき、まず、キーボード１１から文献検索システム起動のコマンドを入力する。これに応じて、検索インタフェイス作動ルーチン４１が起動され、対話的に検索作業を進めるための検索インタフェイス２１が表示手段2 に表示される。図3 は検索インタフェイス２１の初期画面の一例である。検索インタフェイス２１は検索要求入力部２１１、キーワード表示・操作部２１２、ヒット件数表示部２１３、タイトル表示部２１４、文書表示部２１５、検索実行ボタン２１６および特徴語表示ボタン２１７などから構成される。

【０００９】本実施例では、文献検索用のキーワードとして必須キーワード、加点キーワード、減点キーワードの3 種類を用いる。検索は必須キーワードに関するアンドを取って行なわれ、必須キーワードの指定の無い場合には加点キーワードのオアで行われる。必須キーワードで検索された文書が加点キーワードを含む場合には1 点加点し、減点キーワードを含む場合は1 点減点する。同じキーワードが同一文書に何回現れても１点の加点または減点とする。これら3 種類のキーワードに対応してキーワード表示・操作部２１２は3 つの部分から構成される。構成は3 つとも同様であるのでここでは一番左の必須キーワードを主体に説明する。キーワード表示・操作部２１２における必須キーワードの部分は、キーワード表示部２１２１、移動ボタン２１２１１、クリアボタン２１２１２から構成される。移動ボタン２１２１１は他の種類のキーワードを必須キーワードに移す場合に使い、クリアボタン２１２１２はキーワードを必須キーワードから除去する場合に用いる。すなわち、加点キーワードあるいは減点キーワードに表示されているキーワードを選択して必須キーワードの移動ボタン２１２１１を押せば、選択されたキーワードが必須キーワードに移される。また、必須キーワードに表示されているキーワードを選択してクリアボタン２１２１２を押せば、選択されたキーワードが必須キーワードから除去される。同じように、必須キーワードに表示されているキーワードを選択して、たとえば、加点キーワードの移動ボタン２１２２１を押せば、選択されたキーワードが加点キーワードに移される。また、必須キーワードに表示されているキーワードを選択して、減点キーワードの移動ボタン２１２３１を押せば、選択されたキーワードが減点キーワードに移される。

【００１０】また、これらの移動ボタンは後述するように、表示されている特徴語をキーワードにコピーするためのコピーボタンとしても使用される。すなわち、移動かコピーかは対象となる語がどこの領域にあるかにより使い分けられる。検索要求を入力する場合には、検索要求入力部２１１の検索要求入力窓２１１１をマウス１２でクリックするなどして入力待ち状態にしてからキーボード１１を用いて必須キーワード、加点キーワードおよび減点キーワード等の検索要求を入力する。続いて入力

8

完了ボタン２１１２を押すと、入力窓２１１１に入力された文字列が形態素解析ルーチン４２へ渡されて単語列に分割され、さらに除外語データベース６４を参照して、そこに登録されている単語を除去した結果がキーワード格納エリア５１１のデフォールトのキーワード格納エリア５１１１または５１１２（図2）へ格納される。ここではデフォールトのキーワードのタイプは必須キーワードとした。また、それぞれの内容はキーワード表示部２１２１または２１２２にリストの形で表示される。この場合、後述する例からも分かるように、形態素解析ルーチン４２が持つ辞書に応じて単語の分割状態が決まる。

【００１１】ここで、検索実行ボタン２１６を押すと検索ルーチン４３が起動され、検索用インデックスデータベース６２（すなわちある単語がどの文書に含まれているかを示すデータ）を参照して、必須キーワードをアンドで含む文書を検索し、結果として得られ文書識別番号の列が検索結果格納エリア５１２へ格納される。なお検索ルーチン４３は必須キーワードで検索された文書について加点キーワードが含まれている場合には加点キーワードの一つについて1 点加点し、減点キーワードが含まれている場合には減点キーワードの一つについて1 点減点するという作業を行ない、この得点も文書識別番号と合わせて検索結果格納エリア５１２へ格納する。必須キーワードの指定がない場合には、検索ルーチン４３は加点キーワードに関する検索を加点キーワードのオアで行ない、以下同様の仕方で得点を計算する。必須キーワードも加点キーワードもない場合には、検索実行ボタン２１６が押されても検索は行なわない。

【００１２】必須キーワードは、検索に際してはアンドで処理されるから、より厳密に検索結果を絞りこみたいときには不可欠であるが、どちらかといえば、検索結果に漏れがない検索をしたいときには、加点キーワードのみとしてこのオアで検索を行い、この検索結果に入って欲しくない事項を含む可能性がある事項を想定できるときは減点キーワードを設定するのがよい。さらに、検索ルーチン４３は検索結果格納エリア５１２に格納された検索結果から得点分布を計算し、その結果を検索結果得点分布格納エリア５１３に格納する。得点分布とは加点または減点の得点が何点の文書が何件あったかを示すデータである。

【００１３】以下「電子出版」を必須キーワードとする検索要求を入力した場合を例に採り説明する。「電子出版」なる文字列を必須キーワードとして検索要求入力窓２１１１に入力した後、入力完了ボタン２１１２を押す。形態素解析ルーチン４２により「電子出版」は「電子／出版」と分割されて必須キーワード格納エリア５１１１に格納され、さらに必須キーワード表示部２１２１の1 行目と2 行目に分割して表示される。図4 は、この段階で検索実行ボタン２１６を押した場合の検索ワーク

9

エリア５１の状態を示したものである。今の例では必須キーワードが「電子」と「出版」なので、それらが必須キーワード格納エリア５１１１に格納されている。それ以外の加点キーワードあるいは減点キーワードは、検索要求入力窓２１１１に検索者によって付与されなかったので、加点キーワード格納エリア５１１２と減点キーワード格納エリア５１１３は空欄のままである。また検索された文書番号とその得点が検索結果文書番号格納エリア５１２に格納されている。この場合には加点キーワードと減点キーワードがないので得点は全て0である。また得点別に件数をカウントして得られるデータが検索結果得点分布格納エリア５１３に格納されている。この場合得点は0のみでそれが７７件あったことを示している。

【００１４】図5は、この検索結果を表示した検索インタフェイス２１の状態を示したものである。必須キーワード表示部２１２１に必須キーワード「電子」と「出版」が表示され、ヒット件数表示部２１３に検索結果得点分布格納エリア５１３の内容が表示され、タイトル表示部２１４には検索された文書識別番号とそのタイトルが1件1行で適当数表示されている。表示されていない文書識別番号とそのタイトルを知りたいときは、いわゆるスクロールバーによって表示に現れる部分をずらせばよい。表示されたタイトルから本文を読んで見たいものがあれば、該当するタイトルの所をマウスなどで指示すれば本文の内容の一部が文書表示部２１５に表示される。表示されていない部分の文書の内容を知りたいときは、同じように、スクロールバーによって表示に現れる部分をずらせばよい。

【００１５】これで「電子出版」に関する文書が７７件検索されたことになるが、次の段階として、さらに検索を特定の対象に絞り込みたい場合、あるいはそうでなくてもこの７７件の文書にはどのような話題が含まれているかを概観したい場合がある。このような場合には検索インタフェイス２１（図3）上の特徴語表示ボタン２１７を押すと特徴語表示手段作動ルーチン４４が起動され、特徴語表示手段２２が表示画面2に表示される。

【００１６】図6は特徴語表示手段２２の一例の詳細を示したものである。特徴語表示手段２２は操作部２２１、キーワード表示・操作部２２２、ヒット件数表示部２２３、特徴語表示部２２４、パラメーター設定部２２５から構成される。キーワード表示・操作部２２２およびヒット件数表示部２２３は検索インタフェイス２１のキーワード表示・操作部２１２およびヒット件数表示部２１３とそれぞれ連動しており、特徴語表示手段２２上の操作によりこれらの表示内容が変化した場合には自動的に検索インタフェイス２１の方のそれぞれの表示も変化する。しかし逆方向、すなわち検索インタフェイス２１上の操作によりキーワードやヒット件数が変化した場合には自動的には特徴語表示手段２２上には反映されない。これを取り込むには、操作部２２１のリセットボタン２２１４を押すと検索インタフェイス２１側の内容がこちらの特徴語表示手段２２側へコピーされる。なお検索インタフェイス２１上の特徴語表示ボタン２１７を押すことで表示画面2に表示される特徴語表示手段２２の初期画面では検索インタフェイス２１上のキーワードとヒット件数が自動的にコピーされる。今の例の場合、必須キーワード表示部２２２１には「電子」と「出版」が表示され、ヒット件数表示部２２３には「得点0：７７件」が表示されている。

10

【００１７】ここで、操作部２２１の特徴語表示ボタン２２１２を押すと特徴語抽出ルーチン４５が起動され、検索結果格納エリア５１２に格納されたデータから最高得点の文書識別番号を読み込み、それらの文書識別番号に相当する文書の内容を解析して、それらに特徴的に含まれる単語（特徴語）と、それら特徴語間の関連性を解析してグラフにした結果を特徴語表示部２２４に表示する。その過程は以下の説明で詳述する。図7は「電子出版」の例で、グラフ格納エリア５４３（図2）に格納されたデータを示したものである。グラフはノードとリンクからなりそれぞれノード格納エリア５４３１と、リンク格納エリア５４３２とに格納されている。格納されるノードデータは各ノードに表示される特徴語（文字列）とそれを特徴語表示部２２４のどこに表示すべきかを示す座標を中心座標で、さらに文字を表示する領域の横と縦の文字数と表示領域のサイズで構成されている（ただし表示領域のサイズについては、使いやすいようにそれらの1／2の値、すなわち中心から端までのサイズにしてある。）。一方、格納されるリンクデータはグラフ上に表示すべき線分の始点座標と終点座標で構成されている。図では、リンク格納エリア５４３２に格納されている始点座標と終点座標の他に、参考までに、それぞれに対応する文字列のデータを付記したが、実際の装置では、このデータは不要である。図8は、操作部２２１の特徴語表示ボタン２２１２が押されて、特徴語のグラフが表示された状態の特徴語表示手段２２を示した図である。グラフ表示ルーチン４４４が、グラフ格納エリア５４３のデータに従って、特徴語表示部２２４に特徴語とこれを結ぶリンクよりなるグラフを表現する。例えば図7のデータから「コンパクト」は座標（１４９，１３１）を中心として、横方向文字数3、行数2で且横方向で両側に２７、縦方向で上下に１８の矩形の領域を表示域として表示される。この実施例では、座標は特徴語表示部２２４の左上を始点として横方向は右向に、縦方向は下向に取る。また、リンクデータは始点と終点の座標で定義される。リンクデータの1番目は、特徴語「出版」と「電子」との中心座標を結ぶことを意味し、２番目のデータは座標（２０３，１３１）から（３０８，４０）への線分を意味する。これらの語の表示に際しては、それぞれのノードの表示領域には文字表示用に背景

に白色不透明の矩形を表示して、ノードの表示領域では、線分を隠すのがグラフとしては見やすいが、一方、リンクを示すグラフの線とノードの表示領域が重なると、グラフの線が現れないことになり誤解を招くことになりかねない。例えば、図7のデータでは、「デスクトップパブリッシング」と「出版物」を結ぶグラフの線は「ニフティサーブ」の表示領域を通過することになるから、「ニフティサーブ」に白色不透明の矩形をつけると、この部分でグラフの線が線としては表われないことになる。その結果、「ニフティサーブ」と「出版物」とがグラフの線で結ばれ、さらに「ニフティサーブ」と「デスクトップパブリッシング」とがグラフの線で結ばれたように見えることになる。図8では、この対策として、背景に白色不透明の矩形を表示する代わりに、グラフの線の始点及び終点の近傍でのみグラフの線が表示されないようにしてそのノードの表示領域の中に入り込むのを避けるとともに、他の表示領域については通過していることがわかる表示とした。白色不透明の矩形をつけてもグラフの線が隠れないように配置することは大変難しく、特に多数の特徴語をグラフ表示しようとすると見やすいサイズでの表示が不可能となりかねない。

【００１８】パラメータ設定部２２５の特徴語表示設定手段２２５１は特徴語表示部２２４に表示する単語数を調節するためのものであり、設定用つまみ２２５１１を左右に動かして所望の数値に設定する。表示部２２５１２にはその設定値が表示され、特徴語抽出パラメータ格納エリア５２１の抽出語数格納エリア５２１３にその値が格納される。なおこの値は特徴語抽出ルーチン４４１によって利用される。以下では特徴語表示手段２２の特徴語表示ボタン２２１２が押されてから、図7に示したようなグラフデータが作成されるまでの過程を説明する。特徴語表示ボタン２２１２が押されると、計算プログラム保持手段4に格納されている特徴語抽出ルーチン４４１以下共起関係解析ルーチン４４２、グラフ配置ルーチン４４３が順に起動される。特徴語抽出ルーチン４４１は検索ワークエリア５１の検索結果得点分布格納エリア５１３から最高得点とその件数を読み込む。図4に示した「電子」と「出版」の例では最高得点（Ｓ）は0点でありその件数（Ｋ）は７７件である。また特徴語抽出パラメータ格納エリア５２１から走査文書数上限値（Ｍ）５２１１を読み込む。（ここではＭ＝３００とする。）これは検索された文書件数Kが大きい時にすべての文書を解析していると時間がかかるので、一定限度Mを越える場合にはM個のサンプル抽出を行なうためのパラメータである。

【００１９】特徴語抽出ルーチン４４１は、次に、検索結果格納エリア５１２を参照し、得点が最高得点Sと一致するすべての文書識別番号についてそれらの内容を検索対象文書データベース６１から読み込み、形態素解析ルーチン４２を用いて単語分割し、出現するすべての種

類の単語についてそれが出現する文書の数（以下これを文書頻度と呼ぶ）をカウントする。この例では最高得点の件数Kが７７件で、走査文書数上限値Ｍ＝３００以下であったのですべての文書を読み込む。なお、該当文書の形態素解析は、データベース保持手段にゆとりがある場合には、あらかじめ全文書を形態素解析した結果を保持しておき、それを読み込むようにすることも可能である。そうすれば、検索の都度形態素解析をする必要がなくなるので解析時間を大幅に短縮できて有効である。こうして得られる単語とその文書頻度のデータは特徴語抽出ワークエリア５２の中の頻度データ格納エリア５２３に格納される。なお上記で該当文書を形態素解析した結果は後にも使うので、単語分割済み文書格納エリア５２２に格納しておく。

【００２０】図9は「電子出版」の例で頻度データ格納エリア５２３に格納されたデータの一部を示す。各単語ごとのデータは単語名、文書頻度、全体文書頻度、頻度比、頻度クラスの5項目で構成されている。文書頻度は上記作業で検索された文書（この場合７７件）の内の何件のにその単語が出現したかを表す頻度である。また全体文書頻度はキーワードによる検索結果に関係なく、検索対象文書全体で何件の文書に使われているかという頻度である。その情報は単語頻度データベース６３に格納されており、そこから該当する単語の頻度情報を取り出して来たものである。ここで、単語頻度データベース６３は予め検索対象全文書を走査して、出現する全ての単語についてその文書頻度をカウントして作成しておくものとする。頻度比は文書頻度を全体文書頻度で割算した値である。例えば一番最初の「ＲＯＭ」では文書頻度が２１で全体文書頻度が１１８３なので頻度比は２１÷１１８３≒０．０１７である。

【００２１】次に、頻度クラスについて説明する。一般にある文書群に特徴的な語は頻度比の大きさにより判断でき、頻度比が大きいほど特徴度が高いと言える。しかし文書頻度が大きく異なる２つの単語を頻度比で比較するのは危険である。低頻度語の場合には全体頻度が低いのでたまたま頻度比が大きくなる確率が高い。たとえば、図9では、「デスクトップパブリッシング」の頻度比は０．７５となっており、頻度比が大きく特徴度が高いと言えるかと言えば、そうではない。これは文書頻度が3にすぎないのに、全体文書頻度も４でしかないためである。そこで文書頻度が大きく異なる単語同士は比較しないよう、予め文書頻度を適当な幅で区分してクラス分けを行ない各クラスで頻度比が大きいものを特徴語として取る。これによって低頻度語から高頻度語までバランス良く特徴語を抽出することが可能となる。以下頻度クラスの決め方の一例の説明である。特徴語ルーチン４４１は頻度クラス分割数（Ｃ）５２１２を読み込む、これはいくつの頻度クラスに分割するかを示すパラメータであり、使用者が設定する。ここではＣ＝5とする（一

13

般にCは1 以上の整数である）。i 番目の頻度クラスをC［ｉ］として、C［ｉ］に属するための文書頻度がf［ｉ］以上f［ｉ＋1］未満であるとする。ただし最大のクラスについては「ｆ［ｉ＋1］未満」のかわりに「ｆ［ｉ＋1］以下」とする。この頻度閾値f［ｉ］の値の決め方であるが、ここではその一例としてK’を該当文書数として、ｆ［ｉ］＝K’の（ｉ／（Ｃ＋1））乗、とする。（検索された文書数Kが走査文書数上限値Mを越えない場合にはK’＝Kであり、Ｋ＞Mの場合にはK’＝Mである。）今の例ではK’＝７７でC＝5 であるから、ｆ［１］＝７７の（１／6）乗＝2．０６，以下、ｆ［２］＝4．２５，ｆ［３］＝8．７７，ｆ［４］＝１８．１０，ｆ［５］＝３７．３３となる。従って、クラス1：文書頻度3 以上4 以下、クラス2：文書頻度5 以上8 以下、クラス3：文書頻度9 以上１８以下、クラス4：文書頻度１９以上３７以下、クラス5：文書頻度３８以上７７以下、である。

【００２２】この分類条件に従って、各語の文書頻度からそれらの語の頻度クラスを決める。「ＲＯＭ」の場合には文書頻度が２１なのでクラス4 、また「インタラクティブ」は文書頻度が5 なのでクラス2 となる。なお文書頻度がクラス1 よりも小さい場合（この場合文書頻度2 以下）については特徴語抽出の対象から除外する。上記の頻度クラスの付与は次の式で直接計算することもできる。ただしその値がCと一致する場合には1 を引き算する。

（頻度クラス）＝{ｌｏｇ（文書頻度）÷ｌｏｇK’×（Ｃ＋1）}を越えない最大の整数値－1

続いて特徴語抽出ルーチンは抽出語数（ｐ）５２１３を読み込み、各頻度クラスから頻度比が上位のものを合計でこの個数になるように抽出する。それを実現する方法の一例としては、抽出語数p を頻度クラス分割数C で割算して得られる商をｎ、余りをr として、頻度クラスが1 以上r 以下のクラスからはn ＋1 個取り、頻度クラスがr より大きいクラスからはn 個取るという方法がある。

【００２３】以下抽出個数p が１０であるとして図9 の例で説明する。分割数C は5 なのでp ÷Cの商n は2，余りｒ は0 である。従ってクラス1 ～5 から均等に2 個づつ取ることになる。頻度データ格納エリア５２３のデータから各頻度クラスのものについて頻度比が大きいものから順に2 個ずつ取る。図9 のデータより、クラス5 の単語を頻度比が大きい順にならべると「出版」（０．０２７），「電子」（０．０１５），「メディア」（０．００６），「情報」（０．００１）となる。従って上位2 つの「出版」と「電子」が特徴語として取られる。以下同様にしてクラス4 からは「ＲＯＭ」と「コンパクト」、クラス3 からは「メール」と「出版物」、クラス2 からは「インタラクティブ」と「ニフティサーブ」、クラス1 からは「デスクトップパブリッシング」

14

と「パブリッシング」が特徴語として抽出される。それらは特徴語リスト格納エリア５２４に格納される。

【００２４】図１０は特徴語リスト格納エリア５２４に格納されたデータの例である。上記プロセスにより抽出された特徴語とそれらの文書頻度が格納されている。図では、参考に頻度クラスも示したが、これはなくても良い。以上で特徴語抽出ルーチン４４１を抜け、続いて共起関係解析ルーチン４４２が特徴語間の共起データ関係を解析し、結果を共起データ格納エリア５３１に格納する。

【００２５】共起データ格納エリア５３１は特徴語リスト格納エリア５２４に格納された特徴語の集合を縦横に持つ2 次元の配列である。各要素は対応する単語対が共通して現れる文書の数を表す。共起関係解析ルーチン４４２は検索された文書群を単語分割したものを単語分割済み文書格納エリア５２２から読み込み、各文書ごとに共出現するすべての特徴語ペアについて、共起データ格納エリア５３１の対応する要素をインクリメントしていく。

【００２６】次に共起関係解析ルーチン４４２は各特徴語対に対して共起強度を計算する。共起強度は上記作業でカウントされた共起頻度を単語ペアの後者（表では列に当たる単語）の文書頻度で割った値である。単語の文書頻度は特徴語リスト格納エリア５２４に格納されている値（図１０）を用いる。図１１は、この段階における共起データ格納エリア５３１に格納されたデータを示す。各桝目は二つの数値から構成され、上段が対応する単語対の共起頻度、下段が単語対の共起強度（共起頻度÷列側の単語の文書頻度）である。例えば6 行3 列の上段数値6 は、６行目の特徴語「出版物」と３列目の特徴語「ＲＯＭ」が6 件の文書に共出現したことを意味する。この場合単語対の列側の単語「ＲＯＭ」の文書頻度は２１なので、下段の共起強度の数値は6 ÷２１≒0．２９となる。共起データ格納エリア５３１では特徴語は文書頻度の高い順に並べている。後の作業で用いるのは表の対角線の下半分だけなので、残りの部分は省略した。

【００２７】続いて、共起関係解析ルーチン４４２はこの共起データから共起度の高い単語ペア（特徴語グラフでリンクを張るべきペア）を抽出する。本実施例では特徴語間の関連性を示すリンクを、各単語から見てそれより文書頻度が高い単語の中で共起強度の値が最も大きくなる単語に張ることにした。共起関係解析ルーチン４４２はこの基準に従ってリンクを張るべき単語対を集め共起リンク格納エリア５３２に格納する。なお、共起強度が2 番あるは3 番のものでも、１番のものと比べてそれほど小さくない場合（例えば1 番の0．９倍以上）には、リンクを張るというやり方も有力である。図１２はこの段階における共起リンク格納エリア５３２の内容を示す図である。これらのリンクが抽出された過程を図1

15

1の例に基づいて説明をする。図12の2番目の「出版」について見ると、文書頻度が「出版」以上のものは「電子」しかないので「出版」から「電子」にリンクが張られる。次に3番目の「ROM」についてみると、それより頻度が高いのは「出版」と「電子」の2つであり、それらとの共起強度は共に0．27である。この場合には共起データ格納エリア531における番号の小さい「出版」の方にリンクを張る。次に4番の「コンパクト」についてみると、3番の「ROM」との共起強度が0．81で最も大きい。従って「コンパクト」からは「ROM」へリンクを張る。以下同様の操作を続け、図12のようなリンクデータが得られる。

【0028】以上で共起関係解析ルーチン442を抜け、続いて、グラフ配置ルーチン443が起動される。特徴語リスト格納エリア524のデータ（図10）と共起リンク格納エリア532のデータ（図12）にもとづいて特徴語群をノードとするグラフを実際に2次元平面に配置するという作業を行なう。図13はグラフ配置ルーチン443の詳細である。グラフ配置ルーチン443はy座標計算ルーチン4431、x座標計算ルーチン4432、表示座標への変換ルーチン4433、重なり回避ルーチン4434、リンク配置ルーチン4435から構成され、この順に起動する。y座標計算ルーチン4431およびx座標計算ルーチン4,32は表示領域が［－1，1］×［－1，1］の正方形領域であると仮定して各ノードを配置すべき座標を計算する。この座標を正規化された座標と呼ぶ。計算された座標データは正規化座標格納エリア541に格納される。

【0029】初めにy座標計算ルーチン4431が起動され、計算式:

y＝（6／π）×arctan（0．2×log（f／fm））

に従って各特徴語の文書頻度f からそれを表示すべき位置の正規化されたy座標を計算する。すなわち、文書頻度の大きいもの程y軸上では上段に配置されるようにする。ここでf mは特徴語を文書頻度順に並べた時にちょうど真中に来るものの頻度である（ただし偶数個の場合には（個数÷2＋1）番目とする）。実施例では、「電子」「出版」の文書頻度77が最上段となり、「出版物」の文書頻度9が中央位置に当たる。πは円周率、対数log は自然対数、arctanは正接関数の逆関数であり、角度はラジアンを単位とする。例えば「コンパクト」の頻度は21なのでその正規化されたy座標は（6／π）×arctan（0．2×log（21÷9））≒0．32となる。その他の特徴語の正規化されたy座標も同様に計算する。次にx座標計算ルーチン4432が起動され各特徴語表示位置の正規化されたx座標を計算する。図14はx座標計算ルーチン4432の詳細を示した図である。初めにステップ44321により親ノード（リンク先）のないノードが集められる。こ

16

の場合には「電子」のみがそれに当たる。したがってそのx座標の値がステップ44321中の式xi＝－1＋2i／（r＋1）にi＝1を代入して－1＋（2×1）／（1＋1）＝0と計算される。

【0030】続いてループ44322に入り、ステップ44323ではx座標の定まったノード（この場合「電子」のみ）へリンクが張られているノードを一つ取る。共起リンクのデータ（図12）からここでは「出版」がその条件を満たしていることが分かる。続いてステップ44324に入りステップ44323で選ばれたノードの親ノードの集合を求め、さらにそれらのx座標の平均値を計算する。「出版」の親ノードの集合は{「電子」}であり、そのx座標の平均は0である。次にステップ44325では親ノードの集合が{「電子」}と一致するノードを集める。ここではそれは「出版のみである。

【0031】続いて分岐ステップ44326へ入るが親ノードのx座標の平均値が0なのでステップ44327が選択され、「出版」のx座標が計算される。ステップ44327の計算式にs＝1、xp＝0、i＝1を代入して、「出版」のx座標が0と計算される。以上で「電子」と「出版」の正規化されたx座標が定まった。しかしまだ全てのノードのx座標が定まってはいないのでループ44322を繰り返す。ステップ44323ではまだx座標が定まっていないノードの内、リンクが「電子」と「出版」以外には張られていないノードの一つが選択される。この場合「ROM」がその条件を満たす。ステップ44324では「ROM」のリンク先の集合を求め{「出版」}を得る。また親ノード{「出版」}のx座標の平均値xpが0と計算される。

【0032】ステップ44325ではリンク先の集合が{「出版」}と一致するようなノードを集める。「ROM」以外では「メール」がそれに当たる。

【0033】親ノードのx座標の平均値xpが0なので分岐44326では上段が選択され、ステップ44327により「ROM」と「メール」のx座標がそれぞれ［－1，1］を3等分して、－0．33，0．33というように計算される。以下同様にして、すでにx座標が決まったノードのみにリンクが張られるようなノードについて、リンク先が共通のものを集め、親のx座標の平均を中心として区間［－1，1］内に収まるよう均等に配置するようにx座標を決めていく。

【0034】図15は「電子出版」の例でこの段階における正規化座標格納エリア541に格納された座標データを示した図である。つづいて、グラフ配置ルーチン443は表示座標への変換ルーチン4433を起動し、上記の［－1，1］×［－1，1］領域に正規化された座標を特徴語表示部224における実際の位置を表す座標への変換を行ない、ノード格納エリア5431の中心座標欄（図16）に格納する。変換は次のような1次式で

17

行なう。X＝R_x×(1＋x)＋O_x，Y＝R_y×(y m－y)＋O_y。ここで小文字のxとyが正規化された座標、大文字のXとYが特徴語表示部224における座標である。y mはyの最大値を表す。図15の例ではy m＝0．774である。なお係数R_x、R_y、O_x、O_yはグラフ配置パラメータ格納エリア542（図2）の該当するエリアに格納された値を用いる。本例ではR_x＝200，R_y＝200，O_x＝60，O_y＝40とした。上記の一次変換により例えば「コンパクト」の場合、正規化された座標が(－0．555，0．320)なので，X＝200×(1－0．555)＋60＝149，Y＝200×(0．774－0．320)＋40≒131というように計算される。このようにして、全てのノードの特徴語表示部224上での実座標が計算され、ノード格納エリア5431に格納される（図16）。この時次のステップへの準備として単語の順序は、x座標が小さい順に並べる。また文字表示領域の大きさとして横方向の文字数hと行数v、また文字表示領域の横サイズHと縦サイズVを計算して、ノード格納エリア5431に格納する。

【0035】文字表示領域サイズは次の計算式に従って計算する。文字は横書きとし横サイズの限度をW文字とする。Wの値は文字表示部の横方向文字数上限値5426に格納されている値を使う。ここではW＝3とする。表示すべき文字数をMとした場合、横方向の文字数h、と行数vはM≦Wの場合、hはM、vは1である。またM＞Wの場合には、hはWであり、vは(M÷W)以上の最小の整数である。例えば「電子」については文字数が2でこれは横幅限度のW＝3より小さいので、行数vは1で横幅hは2となる。また「インタラクティブ」の場合には文字数が8で横幅限度W＝3を越えるので行数vは(8／3)以上の最小の整数、すなわち3となり、横幅hはW＝3である。また文字表示領域の横サイズの2分の1の値Hと縦サイズの2分の1の値Vはそれぞれの文字数hとvから次の式により計算される。ここで2分の1の値を取ったのは後の処理で主にこの2分の1の値を用いるからである。H＝h×F／2_1＋m_x、V＝v×F／2＋m_y。ここでFは文字フォントの大きさ、m_xはx方向のマージンの大きさ、m_yはy方向のマージンの大きさである。m_xとm_yは2つのノードが接近し過ぎないように、最低限保つべき間隔を表す。F、m_x、m_yはそれぞれ文字サイズ5425、文字表示部の横方向マージン5427、同縦方向マージン5428（図2）に格納されている値を用いる。本例ではF＝16、m_x＝3、m_y＝2とする。例えば「コンパクト」の場合h＝3でv＝2なのでH＝3×16／2＋3＝27、V＝2×16／2＋2＝18と計算される。図16のノード格納エリア5431における文字表示サイズとしての文字数と表示領域サイズはこのようにして計算したものである。

18

【0036】このようにして特徴語表示部における座標が求まったが、この段階ではノードの重なりが生じるおそれがある。例えば図16の例では「電子」と「出版」の座標は同じなので重なってしまう。そのため重なり回避ルーチン4434が起動され、重なりが生じないように座標をずらす操作を行なう。

【0037】図17は重なり回避ルーチン4434の詳細である。全ノードをx座標が小さい順にソートしたものをN[1]，．．．，N[r]とする。N[i]の座標を(X[i]，Y[i])、文字表示領域サイズの値を(H[i]，V[i])とする。i＝2，．．．，rについて次の操作を行なう。j＝1，．．．，i－1の内|Y[j]－Y[i]|＜V[i]＋V[j]となるようなjについてX[j]＋H[j]の最大値を取りξとする。なおそのようなjが無い場合にはこのiについては座標をずらす操作は必要ない。δ＝ξ－(X[i]－H[i])とする。δ≦0の場合にはこのiについては座標をずらす操作は必要ない。δ＞0の場合には、重なりが生じてしまうので、N[i]，．．．，N[r]のx座標をすべて右にδずらす。すなわち、X[k]＝X[k]＋δ(k＝i，．．．，r)とする。

【0038】以上により、全ノードが重ならずに表示できるような座標が与えられる。たとえばi＝2の「インタラクティブ」の場合についてみると、図16のデータより、|Y[2]－Y[1]|＝|240－131|＝109で、V[2]＋V[1]＝26＋18＝44であるから|Y[2]－Y[1]|＜V[2]＋V[1]が成り立たない。従って「インタラクティブ」については横へずらす操作は行なわない。次にi＝3、すなわち「ROM」について見る。j＝1については、|Y[3]－Y[1]|＝|131－131|＝0に対してV[3]＋V[1]＝10＋18＝28となり、|Y[3]－Y[1]|＜V[1]＋V[3]となる。すなわちj＝1の「コンパクト」と重なりが生じてしまう。またj＝2の「インタラクティブ」との関係を見ると、|Y[3]－Y[2]|＝|131－240|＝109、V[3]＋V[2]＝10＋26＝36で|Y[2]－Y[3]|＜V[2]＋V[3]とならないので「インタラクティブ」とは重なる恐れがない。従ってj＝1についてのみx座標を考慮すれば良い。ξ＝X[1]＋H[1]＝149＋27＝176となり、ずらし幅δはδ＝ξ－(X[i]－H[i])＝176－(193－27)＝10である。従ってj＝3，．．．10についてX[j]をすべて＋10する。(X[3]，Y[3])＝(203，131)となり、図7における「ROM」の座標を得る。以下このステップの繰り返しにより図7のノード格納エリア5441と同じデータが得られる。この文字表示領域の重なり回避の操作でも、前述した文字表示領域とグラフの線の重なりはチェックできないし、実際問題として、限られた表示面

19

積では、これを厳密に避けようとすると、適当な大きさの中で、表示のできないことも起こりうるので、実施例では、これについてのチェックはしないこととした。

【００３９】最後にグラフ配置ルーチン４４３はリンク配置ルーチン４４３５を起動する。リンク配置ルーチン４４３５は共起関係解析ワークエリア５３の中の共起リンク格納エリア５３２に格納された共起リンクを張るべき単語ペアに関する情報と、ノードデータ格納エリア５４３１に格納されている各ノードの座標データから特徴語表示部２２４に表示すべき線分のデータ、すなわち始点の座標と終点の座標を作成してリンクデータ格納エリア５４２２に格納する。例えば図１２の共起リンク格納エリア５３２には「ＲＯＭ」から「出版」へのリンクがある。図7のノードデータ格納エリア５４３１に格納されたデータより、「ＲＯＭ」の座標が（２０３，１３１）であり「出版」の座標が（３０８，４０）であることが分かるので、（２０３，１３１）を始点として（３０８，４０）を終点とする線分のデータがリンクデータ格納エリア５４３２に格納される。以上により表示すべきグラフのデータ（図7）が作成された。以下では特徴語表示手段２２の特徴語表示部２２４に表示された特徴語のグラフ表示を参考にして検索作業を進展させる利用形態の例を示す。

【００４０】図8は「電子出版」に関する特徴語表示の例であるが、ここでユーザが仮に表示された語のひとつである「デスクトップパブリッシング」に興味があるとしよう。この場合には、画面上でその単語の所をマウス１２などで指示してから加点キーワードの移動ボタン２２２２２を指示すると「デスクトップパブリッシング」が加点キーワード格納エリア５１１２に格納され、検索インタフェイス２１の加点キーワード表示部２１２２と特徴語表示手段２２の加点キーワード表示部２２２２に表示される。そこで検索インタフェイス２１の検索実行ボタン２１６もしくは特徴語表示手段２２の検索実行ボタン２２１１を押すと加点キーワードに「デスクトップパブリッシング」を加えた形で検索が実行され検索の絞り込みをすることができる。また図8の特徴語表示部２２４に表示された特徴語の中に興味ある単語を発見できなかった場合には特徴語表示数設定手段２２５１を用いて表示語数を増やすことができる。図１８は特徴語表示語数を２０に増やした場合の例である。この場合には図9のデータの例では、このデータから特徴語抽出ルーチン４４１により、２０個の単語が選択されて、図8のケースで説明したと同様に表示される。ここで仮にユーザは「電子出版」における「情報検索」に興味があったとすれば表示されたグラフに「検索」および「情報検索)」という語が表示されているのでそれを利用できる。特徴語表示部の「検索」と「情報検索」をマウスなどでクリックしてから加点キーワードへの移動ボタン２２２２２を押せばこれらの単語が加点用のキーワードと

20

して付け加えられる。これで検索実行ボタン２２１１を押せば検索の絞り込みができる。また検索を絞り込んだ後で特徴語のグラフを見たい場合には特徴語表示ボタン２２１２を押せば良い。それから検索と特徴語のグラフを連続して行なう場合には検索実行＋特徴語表示ボタン２２１３を押せば以上のステップが連続して行なわれる。

【００４１】次に「情報検索」には興味がない場合、あるいは「情報検索」に関する文書には既に目を通してしまい、それ以外の話題に注目したい場合には、減点キーワードを利用する。すでに「検索」と「情報検索」が加点キーワードに加えられている場合には、加点キーワード表示部２２２２に表示されているこれらの単語をマウスなどで指示してから減点キーワードへの移動ボタン２２２３２を押せばこれらの単語が加点キーワードから減点キーワードへ移動する。なお特徴語表示部２２４に表示されている単語を直接減点キーワードとして利用したい場合には、加点キーワードの時と同様に、該当する単語をマウスなどでクリックした後減点キーワードへの移動ボタン２２２３２を押せば良い。すなわち、本実施例では、検索キーワード間では移動ボタンにより移動の操作が行われ、表示された特徴語とキーワード間では移動ボタンにより複写の操作が行われる。

【００４２】「検索」と「情報検索」を減点キーワードへ移動してから検索を実行すると今度はこれらの単語を含む文書の得点が下がり、相対的にこれらを含まない文書の得点が上がるので「電子出版」に関する文書の内、「情報検索」には関係のない文書に注目することが出来る。図１９は特徴語表示様式選択手段２１７１を備え、特徴語をグラフの形で表示したり、リストの形で表示したりすることを選択できる機能を備えた検索インタフェイス２１の一例である。リストでの表示はグラフで表示した場合と比べて、多数の特徴語を表示する為、特徴語相互の関連性を表示できないので関連性に着目した結果の評価ができないという欠点がある反面、スクロールバーを用いることにより、検索結果に出現する多数の特徴語を一覧できるので、ユーザにとって興味と合致する関連語を発見できる可能性が高くなる長所がある。

【００４３】したがって、図１９に示される特徴語表示様式選択手段２１７１を利用して、まず、検索結果をグラフ表示して特徴語の全体像を相互の関連性も含めて概観して、結果を評価し、これにユーザの興味と合致する関連語が十分に表われない場合には、リスト表示を用いて更に細かく探すという二段階の結果評価ができる。さらに、リストを利用した表示から興味のもたれる語が得られたとき、これをキーワードとして利用して、再度検索からやり直すこともできる。図１９の特徴語表示様式選択手段２１７１で「グラフ」を選択すれば、図8あるいは図１８で説明したように、特徴語のグラフ表示がなされる。図１９に示すように、「リスト」を選択すれ

21

ば、図２０に一例を示すように、特徴語表示部２２４には、特徴語がリストの形で表示される。特徴語表示様式選択手段２１７１で「リスト」を選択した場合でも、検索された文書群から特徴語を抽出する方法は前述したグラフ表示の場合と同じである。ただし、リスト表示の場合、図9に示したように頻度を５クラスとするよりは、高、中、低の3クラス程度とする方が見やすいと考えられるので、図２０の表示例では、頻度クラスの分割数は３とした。図２０において、「リスト」の選択に対応して、特徴語表示部２２４には、高頻度特徴語表示部２２４１、中頻度特徴語表示部２２４２および低頻度特徴語表示部２２４３がそれぞれスクロールバー付きの表示枠が設定され、頻度データ格納エリア５２３の特徴語の頻度クラスデータに対応した特徴語が各表示枠内に表示される。各表示枠内での表示順は、たとえば、頻度比の大きさ順にならべるのが良い。これにより、ユーザは、より一般性の高い特徴語から固有名など特殊性の高い特徴語までを一覧でき、幅広い選択肢から興味に合致した単語を検索できる。

【００４４】実施例2

以下、本発明の第2の実施例を図２１に従って説明する。第1の実施例が独立に使用されるコンピュータによる検索装置の構成例であったのに対し、本実施例では、複数のユーザによる検索要求に応えることのできる検索方法を実現するものである。図２１に本実施例の文献検索方法を実現する他の実施例の全体構成を示す。本実施例は、一つのサーバに複数のクライアントが信号伝送回線を介してアクセスし、クライアント毎に検索サービスを受けることのできるものである。サーバは、サーバ自体をクライアントとしても利用することはないのが一般的である。しかし、本実施例では、クライアントからの問題指摘に応じてサーバもクライアントとしても利用する必要がありうることを考慮して、サーバは、実施例1で説明したのと実質的に同じ構成に通信手段7をプラスした検索装置とした。クライアントは実施例1で説明した構成のうち入力手段1、表示手段2、ＣＰＵ３、計算プログラム保持手段4、計算プログラムを動作させるためのワークエリア5およびバス１００のそれぞれに対応するダッシュを付して示した手段、およびサーバとの連係を取るための通信手段7および出力手段8としてのプリンタ８１よりなる。サーバのバス１００にはインタフェイスＩＦ１が、およびクライアントのバス１００にはインタフェイスＩＦ２、ＩＦ３がそれぞれ設けられて、サーバークライアント間を結ぶ回線ＮＥＴ１，ＮＥＴ２で結ばれる。なお、クライアント２についてはバス１００およびインタフェイスＩＦ２のみを図示して他は省略した。

【００４５】クライアント１が文献検索をしようとするとき、まず、入力手段1のキーボード１１から文献検索システム起動のコマンドを入力する。これに応じて、クライアントと側の通信手段7とサーバ側の通信手段7が通信経路ＮＥＴ１を介して連絡を取り、サーバ側の計算プログラム保持手段4の検索インタフェイス作動ルーチン４１がクライアント１側に送信され、クライアント１側で起動される。この結果、表示手段2に対話的に検索作業を進めるための検索インタフェイス２１が表示される。検索インタフェイス２１が表示された後は、クライアント１はこの画面を利用して実施例1で説明したと同様の手順で検索キーとなる語を入力してゆけば良い。なお、クライアント側では検索インタフェイス作動ルーチン４１のコピーを計算プログラム保持手段4に保持しておいて、これを起動するものとしても良い。また、ＷＷＷプラウザなどのハイパーテキスト閲覧インタフェイスを利用して本検索支援サービスが受けられるようにするのも便利である。その場合には、サーバ側には、検索インタフェイス作動ルーチン４１をクライアント側に送信するためのハイパーテキスト（ＨＴ）を用意する。なお、クライアント側では汎用のハイパーテキスト閲覧インタフェイスが利用できる環境にあることを前提とする。

22

【００４６】表示手段2に表示されているハイパーテキスト閲覧インタフェイスのアドレス入力部から、本検索支援サービスが指定するアドレス（すなわちサーバのネットワーク上でのアドレスと検索インタフェイス作動ルーチン４１を送付するためのハイパーテキストＨＴの存在するファイル名など）を指定すると、双方の通信手段を介して指定されたハイパーテキストＨＴが検索インタフェイス作動ルーチン４１を伴ってクライアント側に送られ、送付された検索インタフェイス作動ルーチン４１はクライアント側計算機で起動され、検索インタフェイス２１が表示手段2に表示され利用可能となる。なお、上記では、直接ハイパーテキストＨＴのアドレスを指定したが、ハイパーテキスト閲覧インタフェイスの閲覧部に表示されているハイパーテキストに、本ハイパーテキストＨＴのアドレスがアンカーとして埋め込まれている場合には、そのアンカーの部分をマウスなどでクリックしても同様の動作をさせることができる。

【００４７】クライアント１が入力した検索要求は通信手段7、7と通信経路ＮＥＴ１を介してサーバ側に伝送され、サーバ側で必要な検索と特徴語抽出とグラフ配置計算が実行されて、その結果が再び通信手段7、7の連絡によりクライアント１側に返信され、クライアント１の検索インタフェイス作動ルーチン４１に手渡され、同ルーチンはそのデータに基づいて特徴語グラフを特徴語表示手段２２に表示する。クライアント１はこの検索結果に応じて実施例1で説明したと同様に、さらに必要な検索操作があればこれに応じたデータを入力すれば良い。このデータは再度サーバ側に伝送され、サーバ側で必要な検索が実行されて、その結果が特徴語表示手段２２に表示される。クライアント１は、必要ならプリンタ

23

ー８１によってプリントされた出力を利用することができる。このようにして、クライアント１は、実質的な検索プログラムを持つことなく、サーバ側で実行された結果のみを利用できる。したがって、クライアント１では、ワークエリア５は初期の入力データおよびサーバから伝送されてきた検索結果と特徴語とそのグラフ配置に関するデータ等を保持する能力があれば足りるから、簡易な装置で充実した検索サービスを受けることができる。

【００４８】

【発明の効果】以上、二つのタイプについて説明したように、本発明によれば、ユーザは、より一般性の高い特徴語から固有名など特殊性の高い特徴語までを一覧でき、幅広い選択肢から興味に合致した単語を検索できる。

【図面の簡単な説明】

【図１】本発明の実施例としての独立に使用されるコンピュータによる検索装置の構成例を示すブロック図。

【図２】ワークエリアのデータの割り当て配置の一例を示す図。

【図３】ユーザとコンピュータとの間の検索インタフェイス表示画面の例を示す図。

【図４】検索実行時に検索ワークエリアに格納されるデータの例を示す図。

【図５】図３に示した検索インタフェイス表示画面が検索実行後に検索結果を表示した例を示す図。

【図６】ユーザが検索キーとしての特徴語を付与するための特徴語表示手段起動時の表示画面の例を示す図。

【図７】ユーザから特徴語表示要求があった時に特徴語グラフ格納エリアに格納されるデータの例を示す図。

【図８】検索された文書群における特徴語のグラフ表示の一例を示す図。

【図９】検索された文書群における単語頻度データの一例を示す図。

【図１０】検索された文書群における特徴語リストの一例を示す図。

【図１１】検索された文書群における特徴語間の共起関係を表すデータの一例を示す図。

【図１２】検索された文書群において特に強い共起関係を有する特徴語対のリストの一例を示す図。

【図１３】特徴語のグラフ配置を計算する計算ルーチンの構成の一例を示すパッド図（ＰＡＤ図、Problem Analysis Diagram）。

【図１４】グラフ配置におけるｘ座標計算方法の一例を示すパッド図。

【図１５】検索結果のグラフ表示の際、表示データを正規化された領域に仮想的に配置する際の座標データの一例を示す図。

【図１６】検索結果のグラフ表示の際、表示データの重なり回避を行なう前のグラフの座標の一例を示す図。

24

【図１７】グラフの表示ノードが重なるのを避けるためのルーチンの詳細の一例を示すパッド図。

【図１８】特徴語表示数を２０にした場合の特徴語のグラフ表示の一例を示す図。

【図１９】特徴語表示様式選択手段を備えた検索インタフェイス表示画面の例を示す図。

【図２０】特徴語のリスト表示の表示画面の例を示す図。

【図２１】検索装置の主体がサーバ側に備えられこれに複数のクライアントがアクセスして検索を行う場合の構成例を示すブロック図。

【符合の説明】

１、１：入力手段、１１、１１：キーボード、１２、１２：マウス、１３、１３：ペン入力手段、２、２：表示手段、２１、２１：検索インタフェイス、７、７：通信手段、８：出力手段、８１：プリンタ８１、ＩＦ１、ＩＦ２、ＩＦ３：インタフェイス、ＮＥＴ１，ＮＥＴ２：回線、２１１：検索要求入力部、２１２：キーワード表示・操作部、２１２１：必須キーワード表示部、２１２１１：必須キーワードへの追加ボタン、２１２１２：必須キーワードの消去ボタン、２１２２：加点キーワード表示部、２１２３：減点キーワード表示部、２１３：検索ヒット件数表示部、２１４：タイトル表示部、２１５：文書表示部、２１６：検索実行ボタン、２１６：特徴語表示ボタン、２１７１：特徴語表示様式選択手段、２２：特徴語表示手段、２２１：特徴語表示手段操作部、２２２：特徴語表示手段のキーワード表示・操作部、２２３：特徴語表示手段の検索ヒット件数表示部、２２４：特徴語表示部、２２４１：高頻度特徴語表示部、２２４２：中頻度特徴語表示部、２２４３：高頻度特徴語表示部、２２５：特徴語表示手段のパラメータ設定部、２２５１：特徴語表示語数設定手段、３：計算プログラム実行手段（ＣＰＵ）、４：計算プログラム保持手段、４１：検索インタフェイス作動ルーチン、４２：形態素解析ルーチン、４３：検索ルーチン、４４：特徴語表示手段作動ルーチン、４４１：特徴語抽出ルーチン、４４２：共起関係解析ルーチン、４４３：グラフ配置ルーチン、４４３１：ｙ座標計算ルーチン、４４３２：ｘ座標計算ルーチン、４４３３：表示座標への変換ルーチン、４４３４：重なり回避ルーチン、４４３５：リンク配置ルーチン、４４４：グラフ表示ルーチン、５：ワークエリア、５１：検索ワークエリア、５１１：キーワード格納エリア、５１１１：必須キーワード格納エリア、５１１２：加点キーワード格納エリア、５１１３：減点キーワード格納エリア、５１２：検索結果格納エリア、５１３：検索結果得点分布格納エリア、５２：特徴語抽出ワークエリア、５２１：特徴語抽出パラメータ格納エリア、５２１１：走査文書数上限値格納エリア、５２１２：頻度クラス分割数格納エリア、５２１３：抽出語数格納エリア、５２２：単語分割済み文書格

25

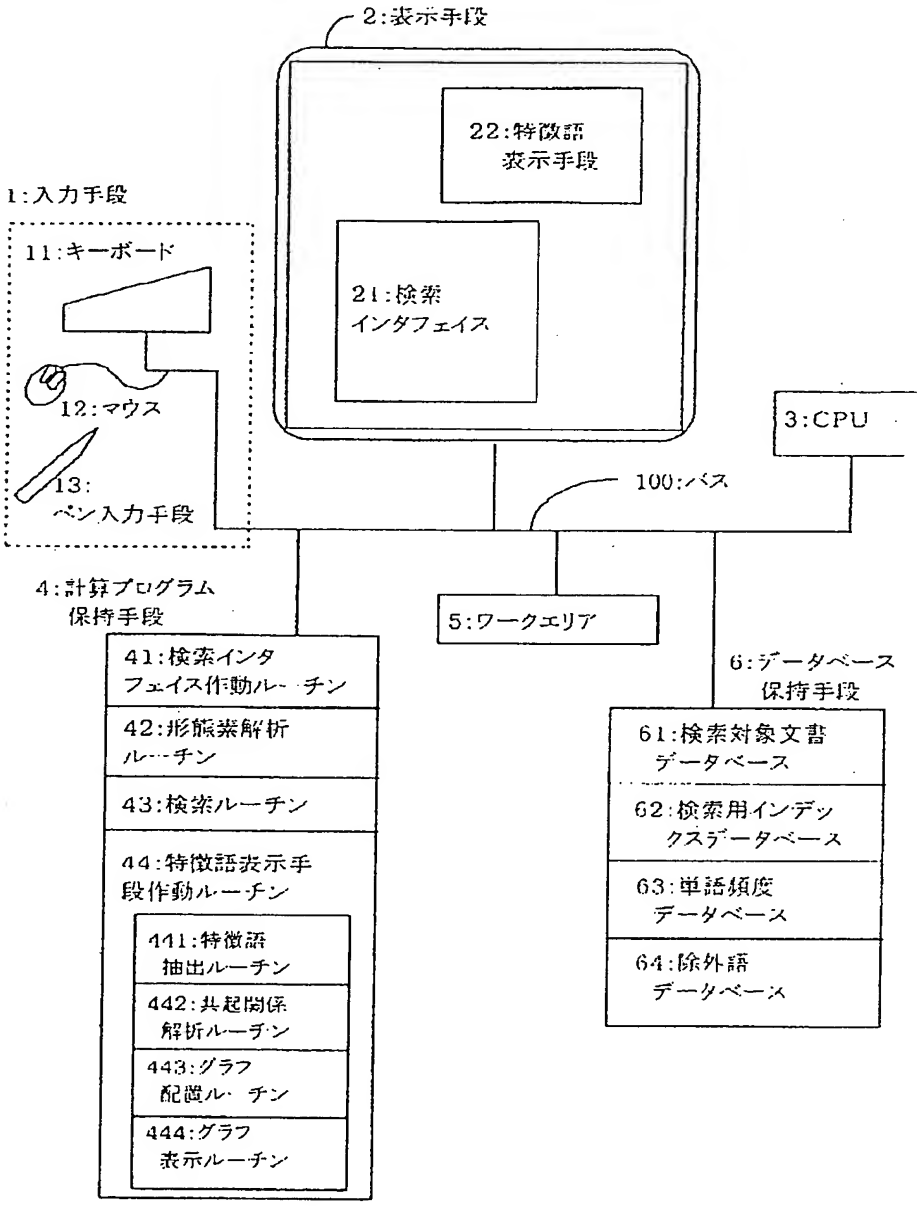
納エリア、５２３：頻度データベース格納エリア、５２４：特徴語リスト格納エリア、５３：共起関係解析ワークエリア、５３１：共起データ格納エリア、５３２：共起リンク格納エリア、５４：グラフ配置ワークエリア、５４１：正規化座標格納エリア、５４２：グラフ配置パラメータ格納エリア、５４３：グラフ格納エリア、５４

26

３１：ノード格納エリア、５４３２：リンク格納エリア、６：データベース保持手段、６１：検索対象文書データベース、６２：検索用インデックスデータベース、６３：単語頻度データベース、６４：除外語データベース。

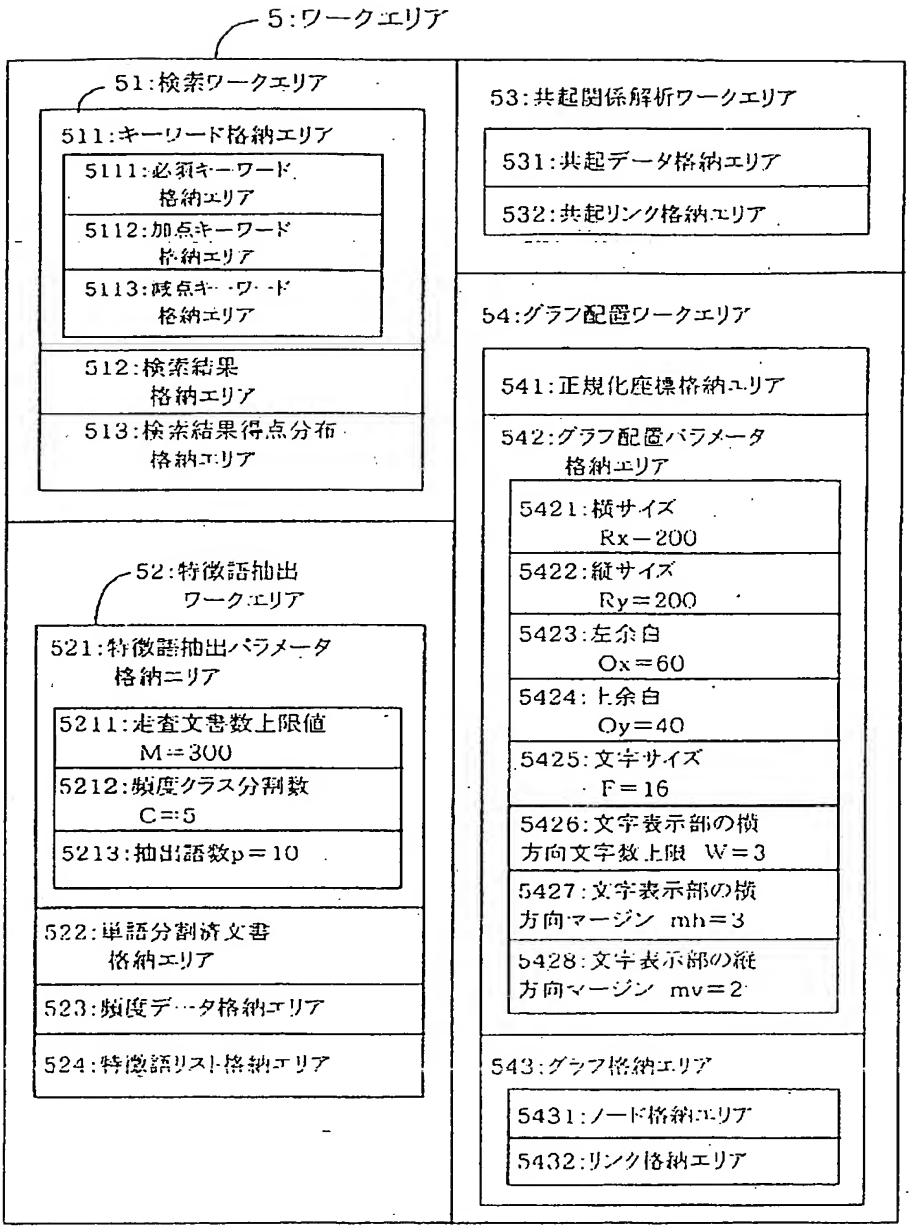
【 図1 】

図 1



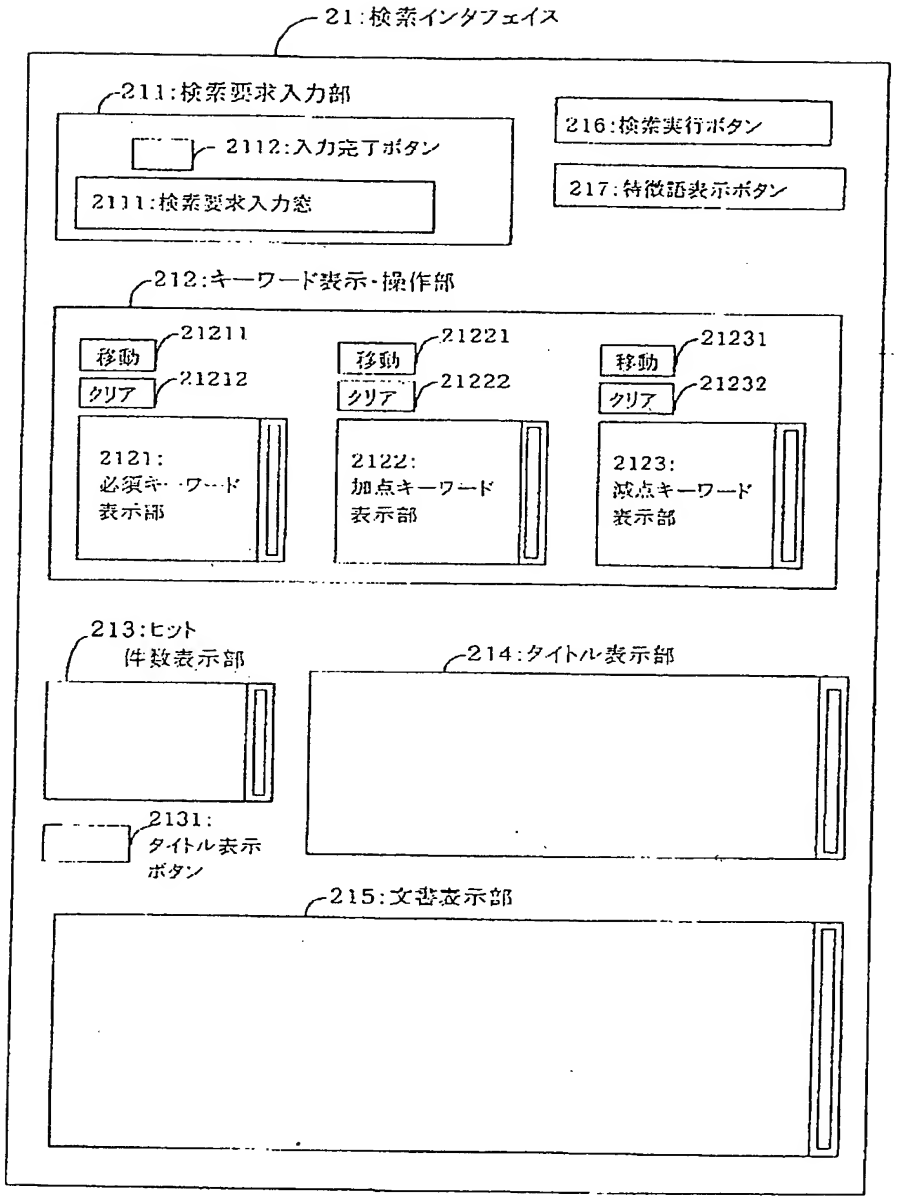
【 図2 】

図 2


5:ワークエリア
51:検索ワークエリア
511:キーワード格納エリア
5111:必須キーワード格納エリア
5112:加点キーワード格納エリア
5113:減点キーワード格納エリア
512:検索結果格納エリア
513:検索結果得点分布格納エリア
52:特徴語抽出ワークエリア
521:特徴語抽出パラメータ格納エリア
5211:走査文書数上限値 M=300
5212:頻度クラス分割数 C=5
5213:抽出語数p＝10
522:単語分割済文書格納エリア
523:頻度データ格納エリア
524:特徴語リスト格納エリア
53:共起関係解析ワークエリア
531:共起データ格納エリア
532:共起リンク格納エリア
54:グラフ配置ワークエリア
541:正規化座標格納エリア
542:グラフ配置パラメータ格納エリア
5421:横サイズ Rx－200
5422:縦サイズ Ry＝200
5423:左余白 Ox＝60
5424:上余白 Oy＝40
5425:文字サイズ F＝16
5426:文字表示部の横方向文字数上限 W＝3
5427:文字表示部の横方向マージン mh＝3
5428:文字表示部の縦方向マージン mv＝2
543:グラフ格納エリア
5431:ノード格納エリア
5432:リンク格納エリア

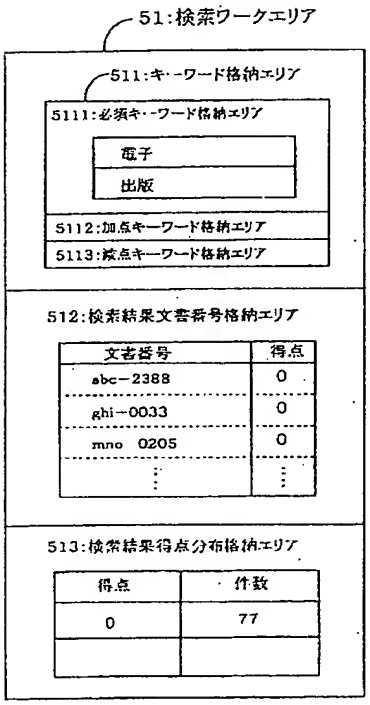
【 図3 】

図 3


21:検索インタフェイス
211:検索要求入力部
2112:入力完了ボタン
2111:検索要求入力窓
216:検索実行ボタン
217:特徴語表示ボタン
212:キーワード表示・操作部
移動
21211
クリア
21212
2121:
必須キーワード
表示部
移動
21221
クリア
21222
2122:
加点キーワード
表示部
移動
21231
クリア
21232
2123:
減点キーワード
表示部
213:ヒット
件数表示部
214:タイトル表示部
2131:
タイトル表示
ボタン
215:文書表示部

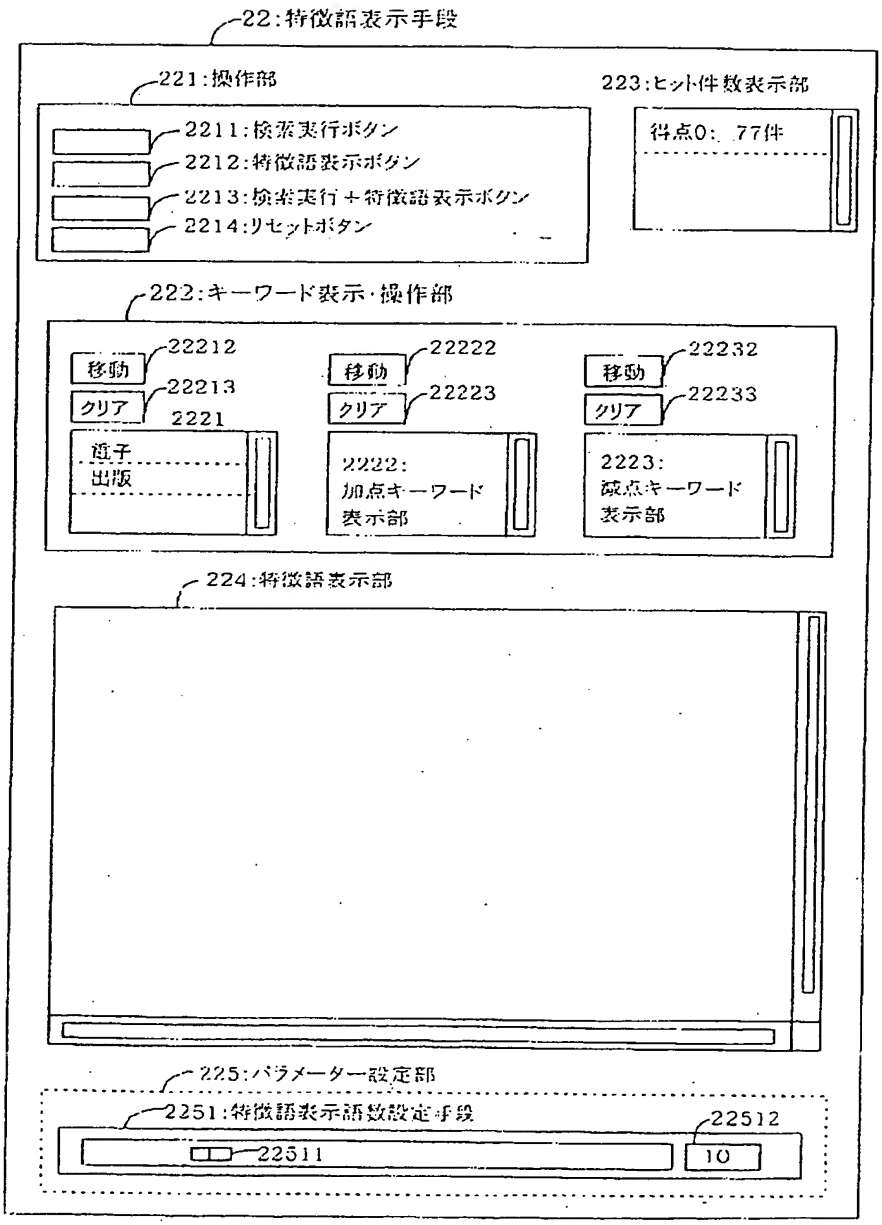
【 図4 】

図 4



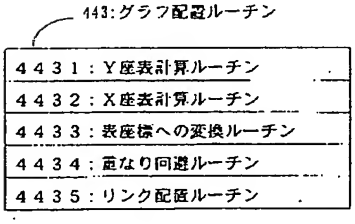
【 図6 】

図 6



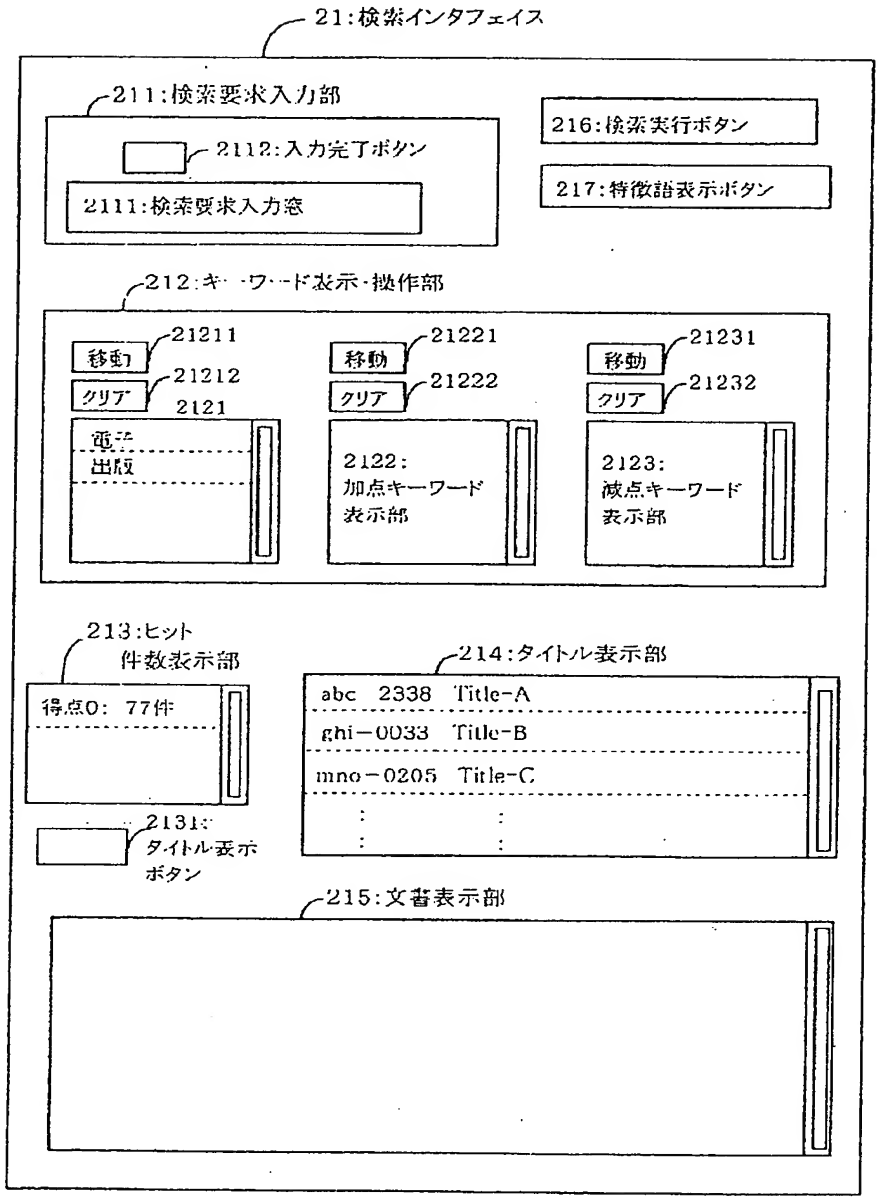
【 図13 】

図 13



【 図5 】

図 5


21:検索インタフェイス
211:検索要求入力部
2112:入力完了ボタン
2111:検索要求入力窓
216:検索実行ボタン
217:特徴語表示ボタン
212:キーワード表示・操作部
21211
移動
21212
クリア
2121
著者
出版
21221
移動
21222
クリア
2122:
加点キーワード
表示部
21231
移動
21232
クリア
2123:
減点キーワード
表示部
213:ヒット
件数表示部
得点0: 77件
2131:
タイトル表示
ボタン
214:タイトル表示部
abc 2338 Title-A
ghi−0033 Title-B
mno−0205 Title-C
215:文書表示部

【図7】

５４３：グラフ格納エリア

５４３１：ノード格納エリア

表示文字列	中心座標 (x, y)	文字数 横(h)	文字数 縦(v)	表示領域サイズ 横(H)	表示領域サイズ 縦(V)
コンパクト	(149, 131)	3	2	27	18
インタラクティブ	(149, 131)	3	3	27	26
ＲＯＭ	(203, 131)	3	1	27	10
出版物	(248, 195)	3	1	27	10
パブリッシング	(248, 256)	3	3	27	26
電子	(270, 40)	2	1	19	10
出版	(308, 40)	2	1	19	10
メール	(375, 161)	3	1	27	10
ニフティサーブ	(375, 240)	3	3	27	26
デスクトップパブリッシング	(429, 278)	3	5	27	42

５４３２：リンク格納エリア

始点座標	終点座標
(308, 40) 出版	(270, 40) 電子
(203, 131) ＲＯＭ	(308, 40) 出版
(375, 161) メール	(308, 40) 出版
(149, 131) コンパクト	(203, 131) ＲＯＭ
(248, 195) 出版物	(203, 131) ＲＯＭ
(375, 240) ニフティサーブ	(375, 161) メール
(149, 240) インタラクティブ	(248, 195) 出版物
(248, 256) パブリッシング	(248, 195) 出版物
(429, 278) デスクトップパブリッシング	(248, 195) 出版物

【図9】

図　９

５２３：頻度データ格納エリア

単語	文書頻度	全体文書頻度	頻度比	頻度クラス
ＲＯＭ	21	1183	0.017	4
ＶＡＮ	6	325	0.018	2
インタラクティブ	5	74	0.067	2
オンライン	13	670	0.019	3
コンパクト	21	978	0.021	4
ディスク	23	1741	0.013	4
ディスクトップパブリッシング	3	4	0.750	1
ニフティサーブ	5	63	0.079	2
ニフティ	6	101	0.059	2
パーソナル	5	215	0.023	2
パブリッシング	4	15	0.266	1
メール	14	524	0.026	3
メディア	41	6821	0.006	5
メモリー	20	1802	0.011	4
印刷物	5	160	0.031	2
検索	15	578	0.025	3
出版物	9	156	0.057	3
出版	77	2800	0.027	5
情報検索	3	42	0.071	1
情報	45	26799	0.001	5
電子	77	4919	0.015	5
電子化	4	58	0.069	1
余白	3	37	0.081	1
・	・	・	・	・
・	・	・	・	・

【図10】

図　１０

５２４：特徴語格納エリア

単語	文書頻度	頻度クラス
出版	７７	5
電子	７７	5
コンパクト	２１	4
ＲＯＭ	２１	4
出版物	9	3
メール	１４	3
ニフティサーブ	5	2
インタラクティブ	5	2
デスクトップパブリッシング	3	1
パブリッシング	4	1

【図11】

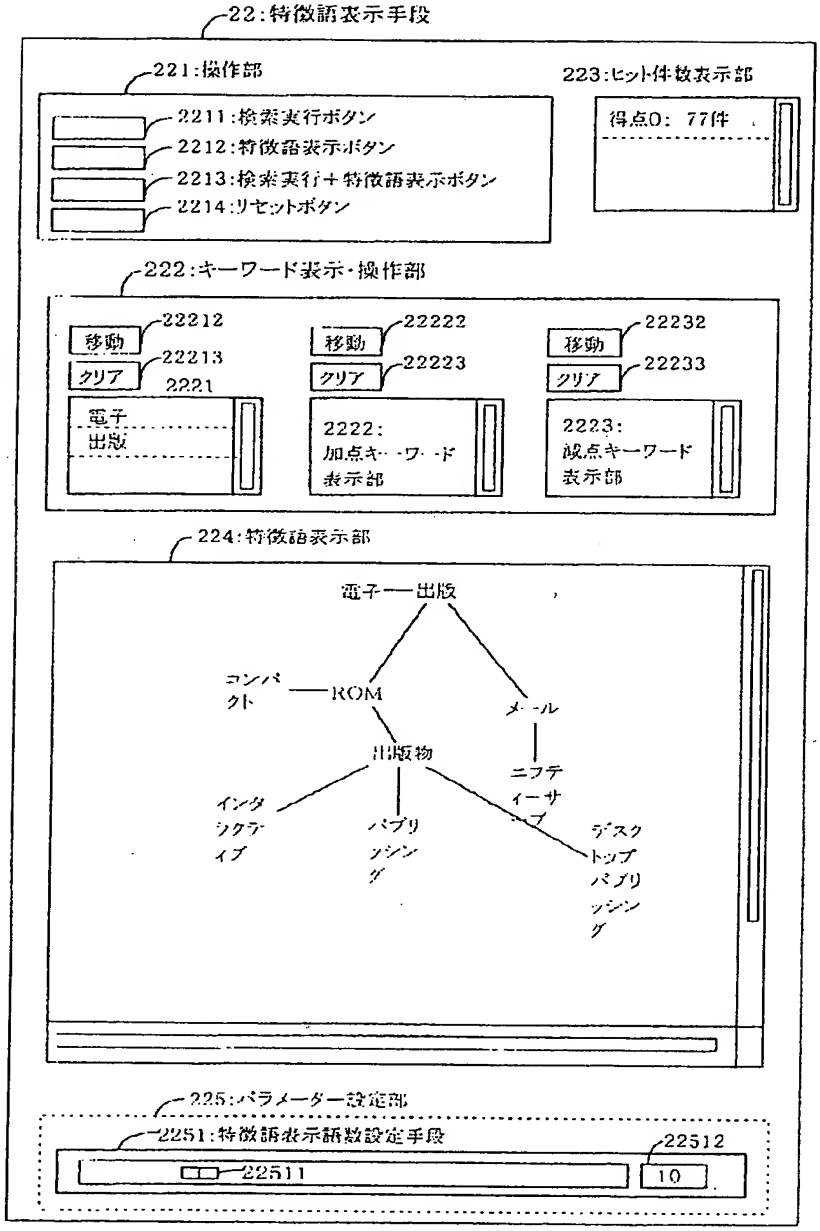
図　１１

５３１：共起データ格納エリア

単語 / 文書頻度	出版 77	電子 77	ＲＯＭ 21	コンパクト 21	メール 14	出版物 9	インタラクティブ 5	ニフティサーブ 5	パブリッシング 4	デスクトップパブリッシング 3
出版	－	－	－	－	－	－	－	－		
電子	77 1.00	－	－	－	－	－	－	－	－	－
ＲＯＭ	21 0.27	21 0.27	－	－	－	－	－	－	－	－
コンパクト	21 0.27	21 0.27	17 0.81	－	－	－	－	－	－	－
メール	14 0.18	14 0.18	1 0.05	0 0.00	－	－	－	－	－	
出版物	9 0.12	9 0.12	6 0.29	5 0.24	1 0.07	－	－	－	－	－
インタラクティブ	5 0.06	5 0.06	2 0.10	3 0.14	1 0.07	2 0.22	－	－	－	－
ニフティサーブ	5 0.06	5 0.06	1 0.05	1 0.05	3 0.21	0 0.00	0 0.00	－	－	－
パブリッシング	4 0.05	4 0.05	0 0.00	1 0.05	1 0.07	2 0.22	1 0.20	0 0.00	－	－
デスクトップパブリッシング	3 0.04	3 0.04	1 0.05	0 0.00	0 0.00	1 0.11	0 0.00	0 0.00	0 0.00	－

【 図8 】

図 8


22:特徴語表示手段
221:操作部
2211:検索実行ボタン
2212:特徴語表示ボタン
2213:検索実行＋特徴語表示ボタン
2214:リセットボタン
223:ヒット件数表示部
得点0: 77件
222:キーワード表示・操作部
移動
22212
クリア
22213
2221
電子
出版
移動
22222
クリア
22223
2222:
加点キーワード
表示部
移動
22232
クリア
22233
2223:
減点キーワード
表示部
224:特徴語表示部
電子出版
コンパクト
ROM
メール
出版物
インタラクティブ
パブリッシング
ニフティーサーブ
デスクトップパブリッシング
225:パラメーター設定部
2251:特徴語表示語数設定手段
22512
22511
10

【 図１２ 】

図　１２

５３２：共起リンク格納エリア

リンク始点	リンク終点
電子	(空)
出版	電子
ＲＯＭ	出版
メール	出版
コンパクト	ＲＯＭ
出版物	ＲＯＭ
ニフティサーブ	メール
インタラクティブ	出版物
パブリッシング	出版物
デスクトップパブリッシング	出版物

【 図１５ 】

図　１５

５４１：正規化座標格納エリア

単語名	Ｘ座標	Ｙ座標
電子	-0.000	0.774
出版	-0.000	0.774
ＲＯＭ	-0.333	0.320
メール	0.333	0.168
コンパクト	-0.555	0.320
出版物	-0.111	0.000
ニフティサーブ	0.333	-0.223
インタラクティブ	-0.555	-0.223
パブリッシング	-0.111	-0.307
デスクトップパブリッシング	0.333	-0.413

【 図１６ 】

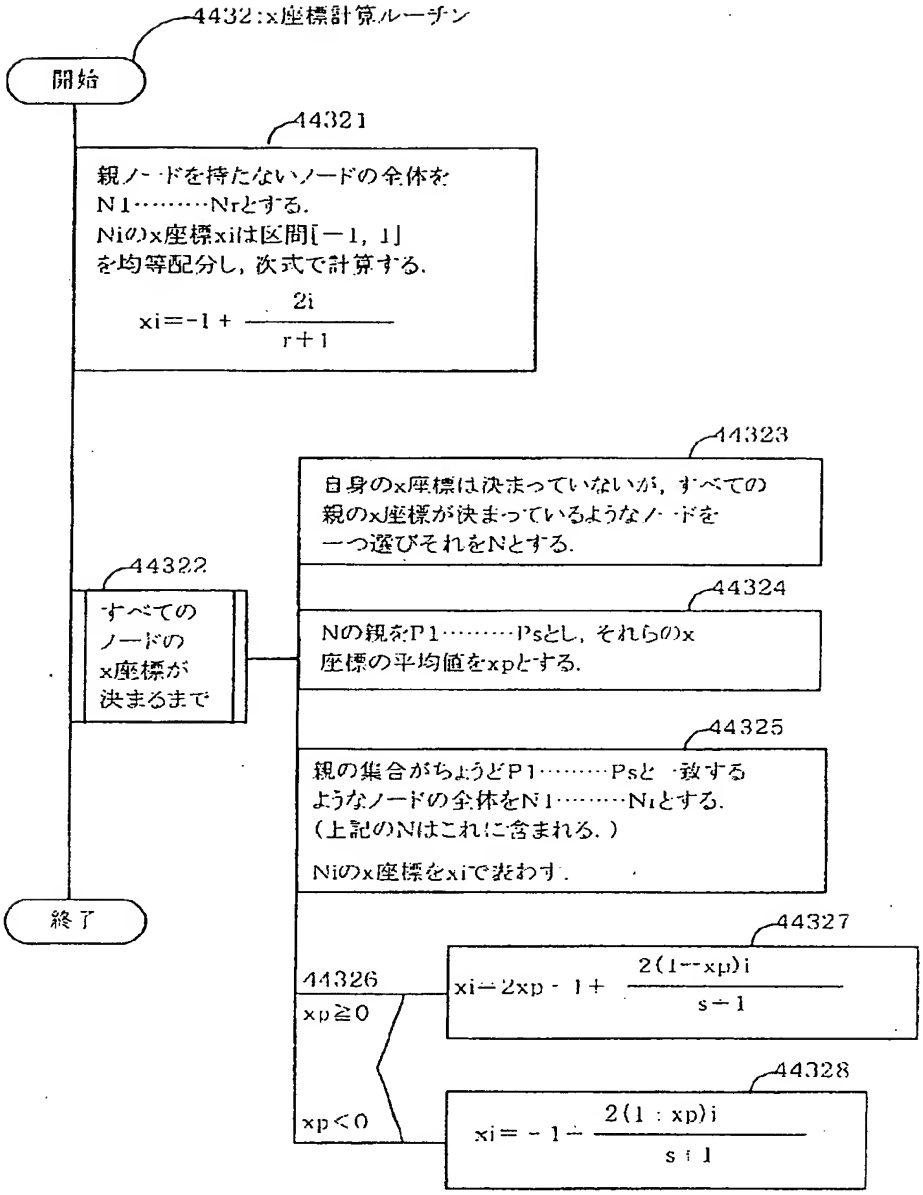
図　１６

５４３１：ノード格納エリア

表示文字列	中心座標	文字数		表示領域サイズ	
	(x, y)	横 (h)	縦 (v)	横 (H)	縦 (V)
コンパクト	(149, 131)	3	2	２７	１８
インタラクティブ	(149, 240)	3	3	２７	２６
ＲＯＭ	(193, 131)	3	1	２７	１０
出版物	(238, 195)	3	1	２７	１０
パブリッシング	(238, 256)	3	3	２７	２６
電子	(260, 40)	2	1	１９	１０
出版	(260, 40)	2	1	１９	１０
メール	(327, 161)	3	1	２７	１０
ニフティサーブ	(327, 240)	3	3	２７	２６
デスクトップパブリッシング	(327, 278)	3	5	２７	４２

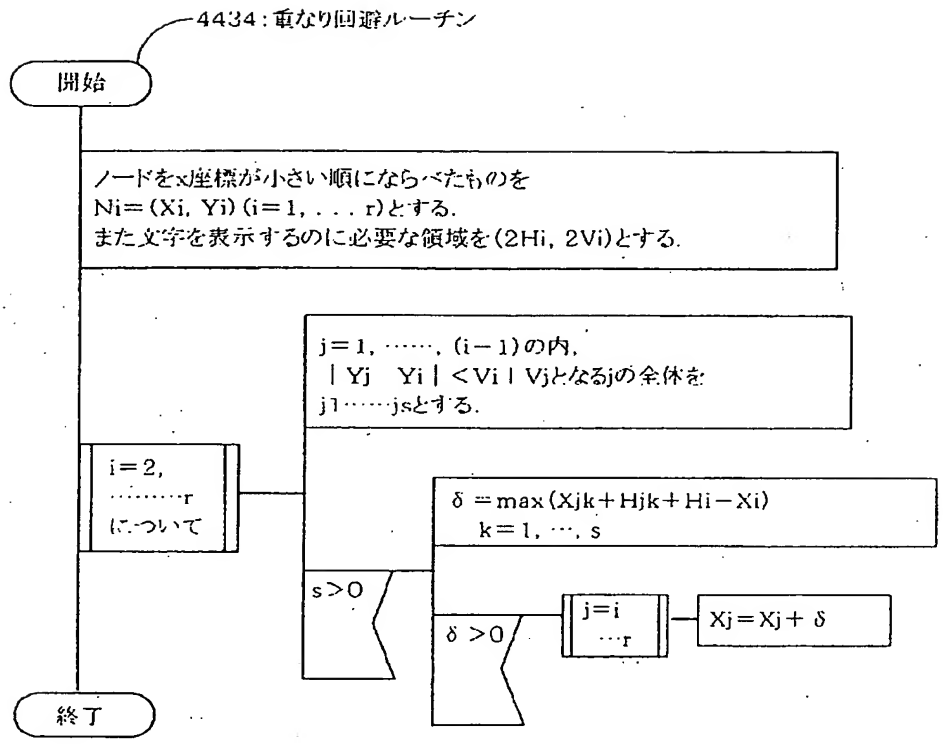
【 図１４ 】

図 14


4432:x座標計算ルーチン
開始
44321
親ノードを持たないノードの全体を
N1……Nrとする.
Niのx座標xiは区間[－1, 1]
を均等配分し, 次式で計算する.
xi=-1+2i/(r+1)
44323
自身のx座標は決まっていないが, すべての
親のx座標が決まっているようなノードを
一つ選びそれをNとする.
44322
すべての
ノードの
x座標が
決まるまで
44324
Nの親をP1……Psとし, それらのx
座標の平均値をxpとする.
44325
親の集合がちょうどP1……Psと一致する
ようなノードの全体をN1……Nrとする.
(上記のNはこれに含まれる.)
Niのx座標をxiで表わす.
終了
44327
44326
xp≧0
xi=2xp-1+2(1-xp)i/(s-1)
xp<0
44328
xi=-1-2(1+xp)i/(s+1)

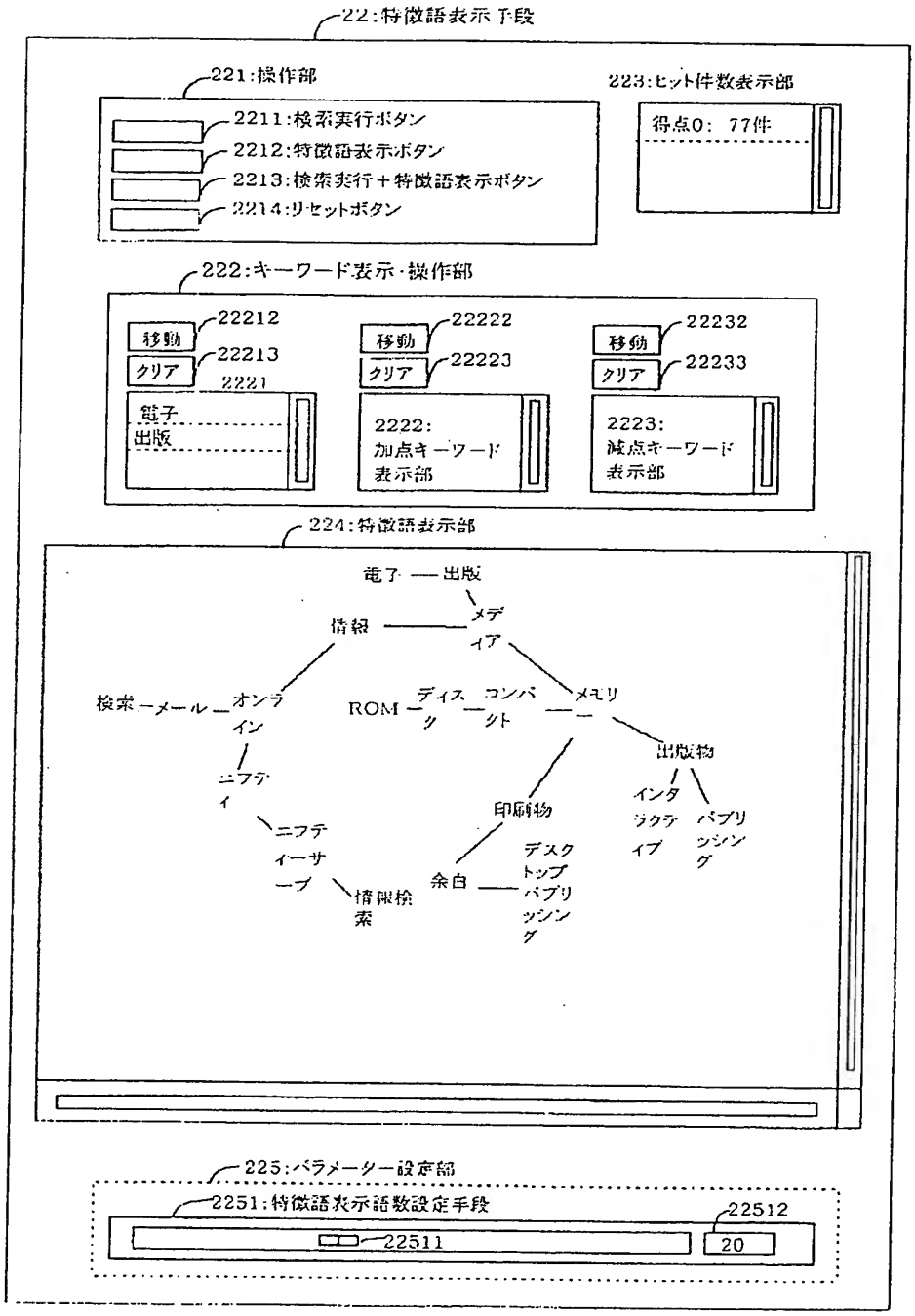
【 図1 7 】

図 17


4434：重なり回避ルーチン
開始
ノードをx座標が小さい順にならべたものを
Ni＝(Xi, Yi) (i＝1, . . . r)とする.
また文字を表示するのに必要な領域を(2Hi, 2Vi)とする.
i＝2,
……r
について
j＝1, ……, (i－1)の内,
｜Yj　Yi｜＜Vi｜Vjとなるjの全体を
j1……jsとする.
δ ＝max (Xjk＋Hjk＋Hi－Xi)
k＝1, …, s
s＞0
δ ＞0
j＝i
…r
Xj＝Xj＋δ
終了

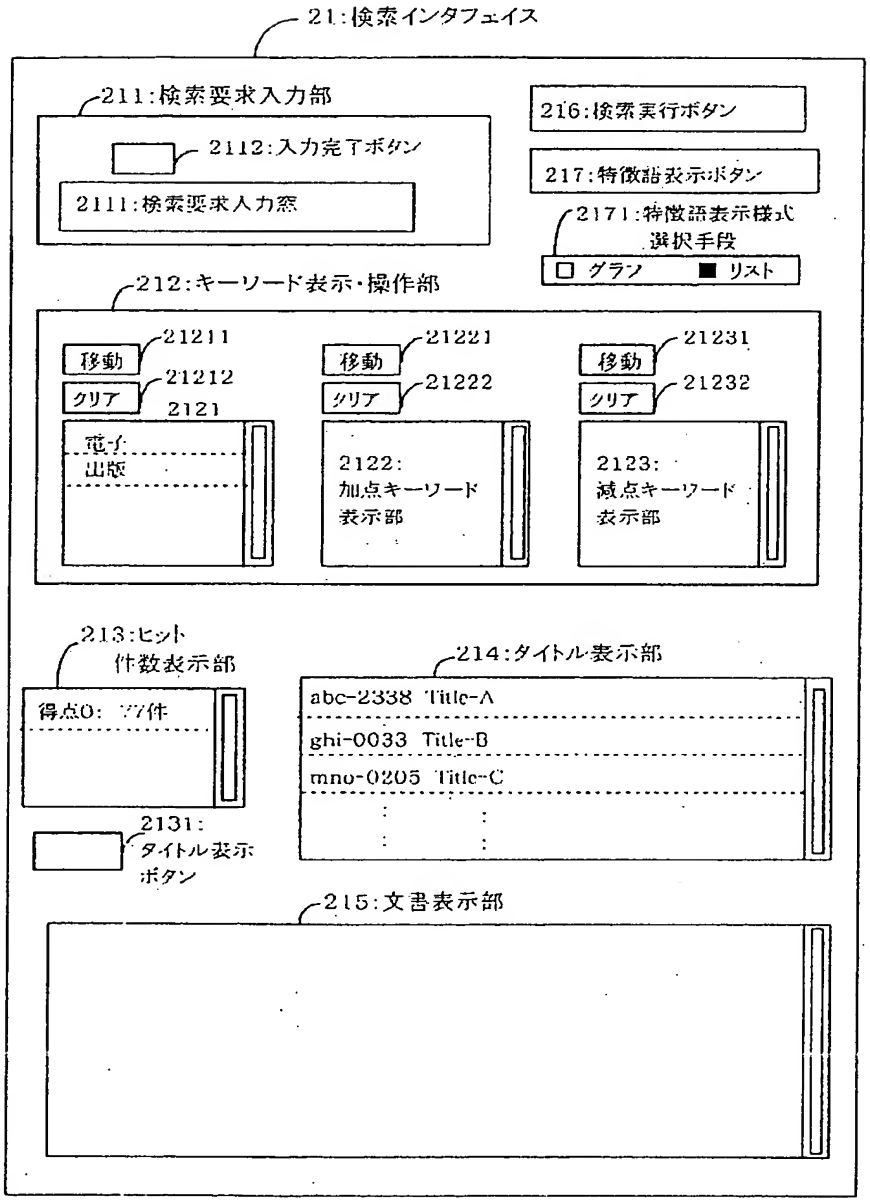
【 図１８ 】

図 18


22:特徴語表示手段
221:操作部
2211:検索実行ボタン
2212:特徴語表示ボタン
2213:検索実行＋特徴語表示ボタン
2214:リセットボタン
223:ヒット件数表示部
得点0: 77件
222:キーワード表示・操作部
移動
22212
クリア
22213
2221
電子出版
移動
22222
クリア
22223
2222:加点キーワード表示部
移動
22232
クリア
22233
2223:減点キーワード表示部
224:特徴語表示部
電子出版
情報
メディア
検索
メール
オンライン
ROM
ディスク
コンパクト
メモリー
出版物
インタラクティブ
パブリッシング
ニフティ
ニフティーサーブ
情報検索
印刷物
余白
デスクトップパブリッシング
225:パラメーター設定部
2251:特徴語表示語数設定手段
22512
22511
20

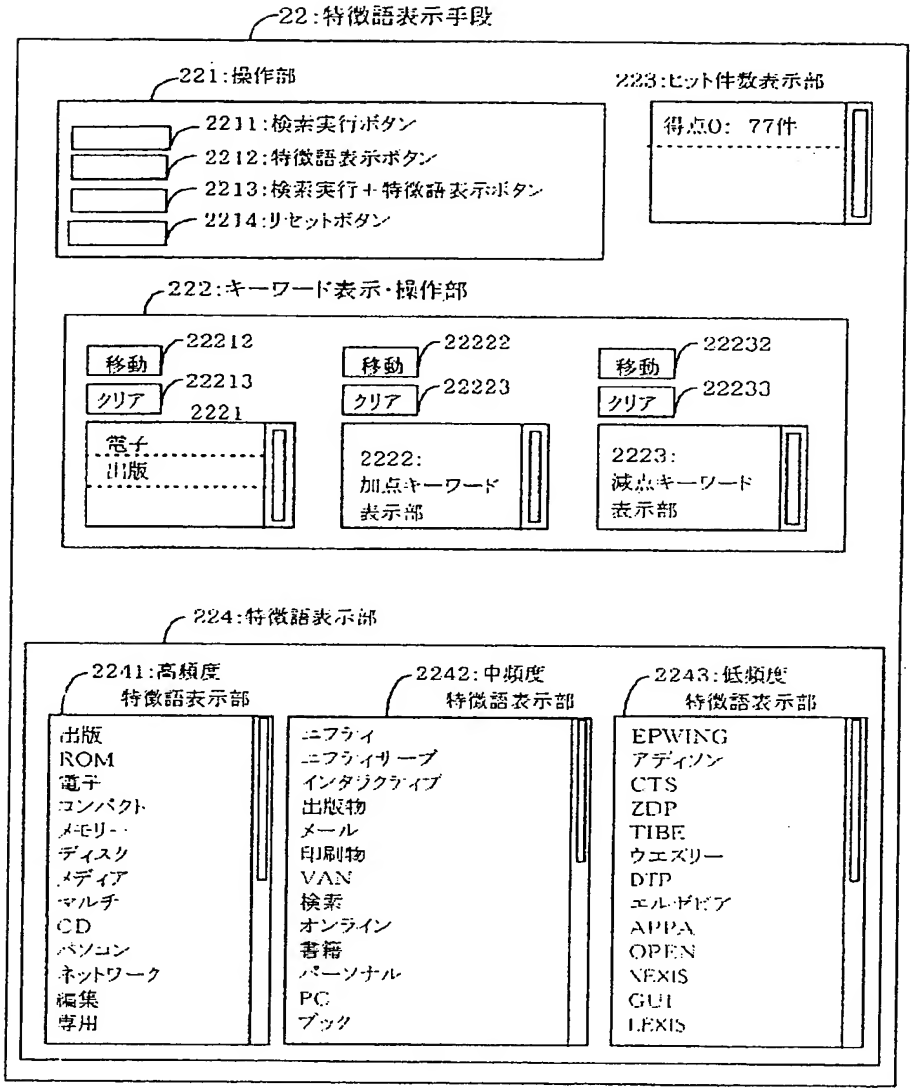
【図１９】

図 19


21:検索インタフェイス
211:検索要求入力部
2112:入力完了ボタン
2111:検索要求入力窓
216:検索実行ボタン
217:特徴語表示ボタン
2171:特徴語表示様式選択手段
グラフ
リスト
212:キーワード表示・操作部
移動
21211
クリア
21212
2121
電子
出版
移動
21221
クリア
21222
2122:
加点キーワード
表示部
移動
21231
クリア
21232
2123:
減点キーワード
表示部
213:ヒット件数表示部
得点0: ??件
2131:
タイトル表示ボタン
214:タイトル表示部
abc-2338 Title-A
ghi-0033 Title-B
mno-0205 Title-C
215:文書表示部

【 図２０ 】

図　20


22:特徴語表示手段
221:操作部
2211:検索実行ボタン
2212:特徴語表示ボタン
2213:検索実行＋特徴語表示ボタン
2214:リセットボタン
223:ヒット件数表示部
得点0: 77件
222:キーワード表示・操作部
移動
22212
クリア
22213
2221
電子
出版
移動
22222
クリア
22223
2222:
加点キーワード
表示部
移動
22232
クリア
22233
2223:
減点キーワード
表示部
224:特徴語表示部
2241:高頻度
特徴語表示部
出版
ROM
電子
コンパクト
メモリー
ディスク
メディア
マルチ
CD
パソコン
ネットワーク
編集
専用
2242:中頻度
特徴語表示部
ユーザ
ユーザサーブ
インタラクティブ
出版物
メール
印刷物
VAN
検索
オンライン
書籍
パーソナル
PC
ブック
2243:低頻度
特徴語表示部
EPWING
アディソン
CTS
ZDP
TIBE
ウエズリー
DTP
エルゼビア
APPA
OPEN
NEXIS
GUI
LEXIS

【 図２１ 】

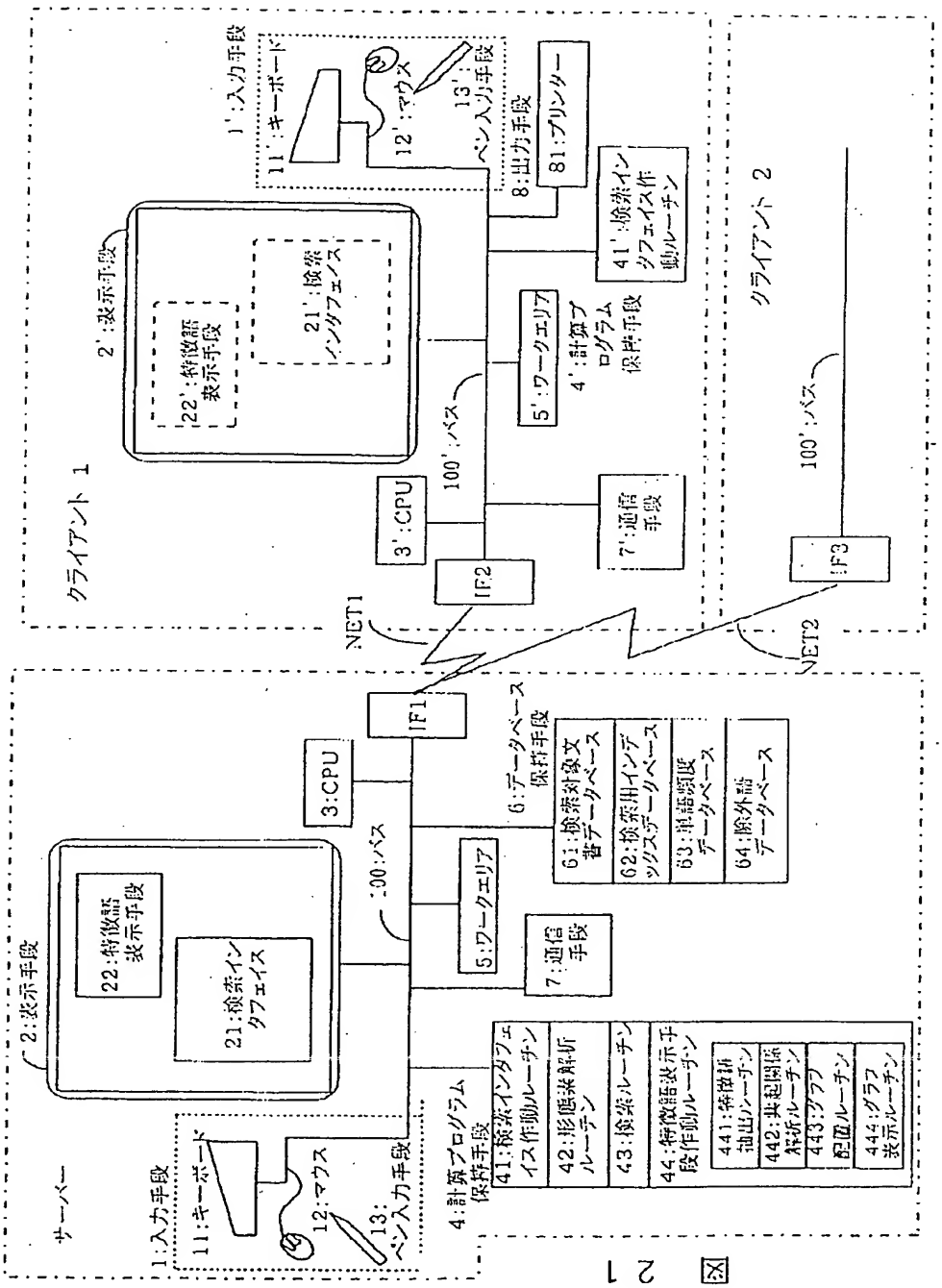

サーバー
1:入力手段
11:キーボード
12:マウス
13:ペン入力手段
2:表示手段
22:特徴語表示手段
21:検索インタフェース
3:CPU
100:バス
IF1
4:計算プログラム保持手段
41:検索インタフェース作動ルーチン
42:形態素解析ルーチン
43:検索ルーチン
44:特徴語表示手段作動ルーチン
441:特徴語抽出ルーチン
442:共起関係解析ルーチン
443:グラフ配置ルーチン
444:グラフ表示ルーチン
5:ワークエリア
6:データベース保持手段
61:検索対象文書データベース
62:検索用インデックスデータベース
63:単語頻度データベース
64:除外語データベース
7:通信手段
NET1
NET2
クライアント 1
1':入力手段
11':キーボード
12':マウス
13':ペン入力手段
2':表示手段
22':特徴語表示手段
21':検索インタフェース
3':CPU
100':バス
IF2
5':ワークエリア
4':計算プログラム保持手段
41':検索インタフェース作動ルーチン
8:出力手段
81:プリンター
7':通信手段
クライアント 2
100':バス
IF3

図 2 1